

VERBAL FACTORS  
MODIFYING SIMILARITY JUDGEMENTS

A thesis presented to the  
Department of Psychology & Sociology  
University of Canterbury

In fulfilment of the requirements for the  
Degree of Doctor of Philosophy

by

Murray Baden Simmonds

March 1971

ACKNOWLEDGEMENTS

The author wishes to thank Professor R.A.M. Gregson for his careful supervision of this thesis. Professor Gregson originally suggested this topic of investigation in the form of an experimental outline of the pilot study reported here. Credit is also due to him for the initial preparation of the computer programs used in experiment 1, for the modification and adaptation of the existing programs used in experiments 2, 3 and 4, and for his patient guidance of the author in devising the model-fitting programs used in experiments 4 and 5.

Thanks are also due to Mr M. J. Mitchell for assistance during Professor Gregson's absence overseas, and to Mr D. O. Watson for advice and assistance in the preparation of the geometric shape stimuli used in experiments 2, 4 and 5, for maintenance of the apparatus used and for advice on the preparation of graphs for this thesis.

P  
37  
S592  
1971

# ABSTRACT

The cognitive processing of sensory information, as exemplified by the judgement of subjective similarity, is discussed in relation to studies of verbal mediation and studies concerned with the role of language in cognition. Five experiments are reported which investigate the effects of pairing labels with stimuli, on the judgement of stimulus similarity. The labels produced small but consistent shifts in stimulus similarity judgements which were shown to be related to the similarity of the labels used. Much of the data were found to be consistent with a trace decay model; there was no evidence that the labels increased the underlying dimensionality of the task. The strategy by which subjects integrated the information present in the stimuli and their accompanying labels appeared to follow an averaging formulation.

## CONTENTS

CHAPTER I.	The Role of Verbal Processes in Cognitive Activities - a review of the literature.	1
	Part I - Some selected studies of verbal mediation.	6
	Part II - The role of language in cognitive processes.	35
CHAPTER II.	An initial investigation of the effects of nonsense labels on taste similarity judgements.	46
	Experiment 1 (Pilot Study).	49
CHAPTER III.	The effect of nonsense labels on the judged similarity of rectangles	89
	Experiment 2.	96
CHAPTER IV.	Some effects of qualitative labels on taste similarity judgements	119
	Experiment 3.	120
CHAPTER V.	Some effects of verbal labels on nonsense shapes.	166
	Experiment 4.	170
	Experiment 5.	209

CHAPTER IV.	A brief summary and directions for future research.	227
	References	235
	Appendix	252

## CHAPTER 1

### THE ROLE OF VERBAL PROCESSES IN COGNITIVE ACTIVITIES

#### - A REVIEW OF THE LITERATURE

The role of verbal processes in conceptual and other complex cognitive activities of the human organism has long been a topic of academic interest. Hull, for example, showed interest in the possibility that language might play a part in such cognitive processes as similarity perception. In a discussion of secondary generalization he concluded that

"..... the commonsense notion of similarity and difference is based upon the presence or absence of primary generalization gradients, whereas so-called logical or abstract similarities and differences arise from secondary, learned, or mediated similarities and differences, particularly those mediated by verbal reactions." (Hull, 1943, p.194.)

It is interesting to note that more than a quarter of a century later, Capehart et al (1969), in a discussion of their new theory of stimulus equivalence (which is based upon information theory and adaptation level theory) pointed out that

"It is recognised that certain equivalence phenomena may be mediated by verbal factors. Under such conditions, A.L. effects may not prevail."

Their avoidance of any attempt to integrate 'verbal factors' with their otherwise general theory of stimulus equivalence is understandable. Since Hull's book was published, there has been very little direct investigation of the role of language in human judgement processes.

The suggestion that similarity judgements may provide a useful vehicle for the investigation of cognitive processes has been made by Wallach (1958, p.104):

"..... the question of psychological similarity may lie at the basis of diverse kinds of psychological research on cognition that have hitherto usually been considered apart from each other - ranging from studies on stimulus generalization to studies on learning and thinking."

Throughout this thesis we will be using the judgement of subjective similarity as our approach to the cognitive mechanisms in which verbal processes are thought to play a part. This approach has been outlined by Høijer (1969, 1970) who is of the opinion that

"any psychological theory of similarity must be a cognitive theory" (1970, p.1).

Similarity will here be viewed as a relation, where a relation can be defined as a set of ordered pairs (Berlyne, 1965). As Reese (1968) has suggested, an ordered pair is a pair of elements presented in a particular spatial or temporal order, symbolised  $a R b$  to denote that 'a' stands in the relation R to 'b'. In

other words, similarity, as one kind of relation, is a variable characterising pairs of objects. Sjöberg (1969) claims that the possibility of measuring similarity has seldom been questioned, although little (data) bearing directly on the issue (has) been published. Beals, Krantz and Tversky (1968) point out that similarity, or psychological distance, is widely used in psychology both as a descriptive and as an explanatory concept. They add that since judgements of the similarity between stimuli contain information about the ways the stimuli are perceived and coded, similarity judgements provide a useful tool in the study of perception and cognition. Yoshida and Saito (in press) and Beals et al (1968) have distinguished the direct and the indirect methods of similarity estimation. Since we are defining similarity in terms of cognitive processes we will be using direct methods, that is, we believe it is possible to accept numerical (or equivalent) judgements from subjects who have been given a suitable scale and adequate instructions to judge similarity. We will assume that, under the conditions employed in this research, subjects behave as rational judges in that intra-individual consistency in scale usage is high. In some cases we will also assume, as an approximation, that inter-individual differences in judgement are normally distributed and that pooling of data from subjects within a given experimental condition is justified.

In summary then, we will define similarity as the ordering of pairs of stimuli that is produced by subjects who respond rationally under instructions to "judge similarity" given an



adequate category scale. This approach to similarity is contrasted with the indirect approach, as exemplified by the Gibson (1940) definition of similarity as "that relationship between stimulus items which can be indicated and measured in terms of their tendency to generalize." The emphasis in this thesis will be on the cognitive processes to which similarity is considered to provide a means of approach, rather than on the nature of the similarity process itself.

The two topics of interest here, cognitive processes as exemplified by similarity judgement, and verbal mediating processes, are certainly not new to psychology; in fact both topics have generated a considerable body of research. However, perhaps largely for historical reasons, the theories and data relevant to each area have tended to develop in divergent directions. It will therefore be necessary to review the literature relevant to the two topics separately, before attempting to postulate any specific relationships between verbal mediation and cognitive processes including the judgement of similarities. The first section of this review, then, will be mainly concerned with the concept of mediation, with the emphasis on a cross-section of the literature that is relevant to the study of verbal mediating stimuli and responses. Part Two of the review will be concerned with investigations of the influences of language on cognition, and will emphasise the ways in which this field differs, in background and in current theoretical and methodological approach, from the research on verbal mediation. Finally in this chapter, attempts that have been made to bridge

these two fields will be outlined and suggestions for an experimental investigation of the role of verbal mediation in similarity judgement will be explored.

PART I.

## SOME SELECTED STUDIES OF VERBAL MEDIATION.

The historical development of the concept of verbal mediation can be sketched in terms of a few key experiments. These will first be described; we will then attempt to gain a more detailed understanding of the term mediation as it is used in contemporary psychology.

Historically, the British Associationists appear to have been the first psychologists to study the problem of mediation. The classic paradigm of mediate association as set forth by them is deceptively simple. Rephrased in modern terminology, it states simply that

"If element A is associated with element B, and element C is also associated with element B, then element A will acquire some association with element C".

In the early 1930's Shipley apparently rediscovered mediation for modern psychology, while performing some experiments on classical conditioning. He put two groups of 10 subjects through an experimental programme consisting of two discrete conditioning steps. The first step consisted of repeated pairings of "extra stimuli" with a sudden rap on the cheek which served to evoke an eyeblink from the subject. In the first group of 10 subjects a buzz and a flash of light were used as the "extra stimuli"; in the second group, the buzz alone was used. The second step in the experiment for both groups consisted of the repeated pairing of the flash of light with a shock, which served to evoke finger

withdrawal. At the end of the experiment, the "extra stimulus" of the first step (buzz) which had never been given in the second step, was presented alone, and was found to evoke finger withdrawals in 60% and 70% of the subjects in each group respectively. An additional group of 10 subjects showed that finger withdrawal did not occur at the test point unless the test stimulus was one that had previously been used in one of the two steps of the experiment. Finally, a further group of 10 subjects showed that finger withdrawal occurred in but a small percentage of subjects (20%) when the shock used in the second step had not been paired with an "extra stimulus". Thus, the reaction did not occur with any appreciable degree of regularity except when the test or critical stimulus had been coupled with some response and the critical response had been coupled with some extra stimulus. Shipley (1935) concluded that the reaction depends on the previous conditioning of both the test stimulus and the critical response, and called it an indirect conditioning reaction.

The experiment suggests that a response, in this case the eyeblink, can in a manner of speaking, "carry" another response (finger withdrawal) from one stimulus (tap on cheek) to another (light flash) without the latter having ever been used, either as a conditioned or an unconditioned stimulus for the elicitation of the finger withdrawal response.

Lumsdaine (1939) raised the question of whether or not the mediating reaction actually occurs. He repeated Shipley's experiment with detailed graphic recording and found that in most

cases the light did elicit a winking reaction which was closely followed by finger withdrawal. There were some cases, however, in which the withdrawal reaction antedated the eye movement and this suggests that the winking reaction may be only an overt index of the actual mediation process. In fact, the mediation hypothesis leads us to expect just this; for in the original training, the light (sign) was presumably becoming associated with the fractional anticipatory portions of the reaction to the tap on the cheek (the stimulus object) and it is this mediation process which is more or less faithfully indexed by the overt winking.

An early extension of the mediation paradigm to verbal behaviour appeared in some studies by Razran (1949). He showed that a salivary response which has been conditioned to one word, e.g. "style" will readily generalise to another word, e.g. "fashion" which is physically very different, but which elicits much the same meaning or mediating reaction. In fact, Razran found that considerably more generalization of conditioning (59% as opposed to 37%) occurred on the basis of similarity of meaning (synonyms) than on the basis of sight or sound (homophones; e.g. for "style" a homophone would be "stile").

To take the mediation paradigm one step further into the field of psycholinguistics, a purely verbal experimental test for mediate association was provided by the following experiment performed by Sacks and Russell (1953). This experiment made use of naturally associated pairs of words, e.g. "umbrella" and "rain". Sacks and Russell had subjects associate one member of the

"umbrella" - "rain" word pair with a nonsense syllable (e.g. zup) by the method of paired associates. They then tested for association between the nonsense syllable and the other member of the natural pair by requiring the subject to learn a new list which contained this pair. His performance was contrasted against other subjects in a control group who had not learned the appropriate first list. Thus, given (from norms)  $A \rightarrow B$  (e.g. Umbrella - Rain)  
 Subject first learns  $C \rightarrow A$  (e.g. Zup - Umbrella)  
 And is then tested with  $C \rightarrow B$  (e.g. Zup - rain).

Under these conditions, enormous facilitation of second list learning was found, in that subjects virtually learned the second list in one trial.

To answer the criticism that subjects were "aware" of the relationship between the words in the first and second list, Russell and Storms (1955) extended the chain by another inferred link:

e.g. Given (from norms)  $\text{soldier} \rightarrow \text{sailor}$ ,  $\text{sailor} \rightarrow \text{navy}$ ,  
 but  $\text{soldier} \nrightarrow \text{navy}$

Subject first learns  $\text{zug} \rightarrow \text{soldier}$   
 and is tested with  $\text{zug} \rightarrow \text{navy}$ .

Again significant facilitation was found in the second list. Note that the assumed implicit stimulus (in this case, the word "sailor") did not appear anywhere in the experiment, and subjects were apparently unaware of the relationship between the first and second lists.

More recent work, e.g. Mink (1957) has shown that mediation

is not necessarily a neat, automatic associative process. Further complications have arisen from the suggestions made by Jenkins (1961) that reinforcement of specific mediation responses made by a given subject may be crucial, and also that there may be some importance attached to the general likelihood of eliciting the mediational processes, as determined either by the characteristics of the tasks confronting the subject, or by his history of reinforcement for mediating in the specific situation. Mandler (1961) however, takes the view that the strategies used by subjects are crucial and claims that a term such as "decision making" may be more appropriate than the term reinforcement. He draws a comparison between some of the more complex mediation paradigms and the typical concept learning experiments in which the subject's task is to acquire and use a strategy; the strategy in this case being the use of a verbal mediating response. Russell (1961) has also pointed out that mediation processes are characterised by a high degree of selective use, and may be used only under the operation of appropriate "sets". Recent research thus appears to have elevated the mediation concept from Shipley's view of it as a relatively simple form of conditioned behaviour to a modern view which places it on a par with other complex cognitive processes like concept formation and judgement. We will return to this point again in Chapter 6. For the present, however, we will be concerned with a more general summary of the literature on mediation processes.

A considerable body of research has been performed using paired associate and related learning tasks to explore various aspects

of the mediation paradigm, and much of it has been reviewed by Cofer and Musgrave (1961). Our aim here, however, is to explore very generally some aspects of the wide variety of research that has drawn upon the concept of mediation, in order to gain some insight into the nature of the process. The problem that we wish to investigate involves the possibility that language mediates in some way between the raw sensory input and the internal or covert reactions to the input which determine the outcome of the complex cognitive activities that occur during such processes as judgement. Areas of research in which mediation and related phenomena are considered to operate include the following:

1. Investigations of mediation in paired associates learning.
2. Investigations of stimulus predifferentiation.
3. Studies of verbal factors in the acquisition of motor skills.
4. Studies of transposition and the perception of stimulus relations.
5. Experiments involving mediation in concept learning.
6. Mediated generalization studies.

Some typical studies from each area will now be discussed as a means of illustrating various aspects of the mediation phenomenon.

#### 1. Mediation in paired associates learning.

Studies relevant to this area have already been discussed in sufficient detail. There have been attempts to show that



mediation operates without awareness of the process on the part of the subject e.g. Russell and Storms (1955). One notable feature appears to emerge however; for as soon as the task is made sufficiently complex for us to be sure that the subject could not be aware of mediation, other factors such as the use of strategies are reported to become significant e.g. Jenkins (1961). The use of these strategies in turn suggests that awareness of the phenomenon is important for subjects to show mediation. In fact, this controversy suggests interesting parallels with that of the problem of awareness in verbal conditioning, (Greenspoon, 1955). Leftwich, Nawas and Siegel (1969) recently investigated the problem and came to the conclusion that cognitive theory, and its construct of awareness provided a better account of the verbal conditioning phenomenon than reinforcement theory.

## 2. Studies of stimulus predifferentiation.

The term stimulus predifferentiation has been referred to the observation that the learning of different responses to physically similar stimuli is reputedly facilitated by the prior pairing of discriminial verbal responses with each member of the set of physically similar stimuli. The responses whose acquisition is reported to be facilitated are usually motor transfer tasks. The typical explanation has been that when verbal responses are learned to a set of physical stimuli, the response-produced stimulation which accompanies the verbalisation contributes to each total stimulus complex; these differential correlates of verbalisation

then enhance the perceived difference between the otherwise similar physical stimuli and thus enhance discrimination and therefore learning (Spiker, 1956). Initial investigations of pre-differentiation were concerned with purely verbal learning problems. The original predifferentiation hypothesis as stated by Gibson, was as follows:

"If differentiation has been set up within a list (of paired associate items), less generalization will occur in learning a new list which includes the same stimulus items paired with different responses; and the trials required to learn the new list will tend to be reduced by a reduction of the internal generalization." (Gibson, 1940)

In its original form, the concept of predifferentiation was thus considered to be purely mediational. Miller and Dollard (1945) extended the concept to motor behaviour by postulating that the verbal responses, which are learned during pretraining, transfer to the motor task and it is this transfer of additional distinctive cues to the discriminanda which reduces the generalization between them, thus facilitating learning of the appropriate motor task.

An experiment by Norcross (1958) is a typical example of the approach, and has sometimes been cited as evidence for the acquired distinctiveness of stimuli, produced by labelling. Her stimuli were line drawings of Indian children's faces; two boys and two girls. Using a fully counterbalanced design, she taught different groups of kindergarten children to use similar names

('zim' and 'zam') for one pair of faces and dissimilar names ('wug' and 'kos') for the other pair. The transfer task consisted of learning to press a different one of 4 buttons to each face. It was predicted that learning performance would be better for the pair of stimuli with dissimilar names; the prediction was confirmed. She concluded that response-produced verbal cues may differentially affect a transfer task, depending upon the degree of generalization among the verbal cue components. Carroll (1964) has put forward a valid criticism of this interpretation. He points out that since her procedure called for the child to verbalise the correct name for each face before he made his motor response, the experiment may be regarded as showing merely that it is harder to learn associations to relatively similar verbal stimuli than to dissimilar verbal stimuli, a finding which has been repeatedly demonstrated in paired associate experiments. In fact, as Carroll suggests, it is difficult to conceive of an experiment that would demonstrate acquired distinctiveness without being subject to the criticism that the discrimination is made in response to words rather than to characteristics of stimuli that are somehow invested in them by the words assigned to them. The hypothesis that discriminatory responses can be made to verbal stimuli is interesting in itself and seems to be a more reasonable interpretation of the experimental results.

A related study by Spiker (1959) is also of interest for the implications it may have for our understanding of the process of mediation. Briefly he showed that preliminary training with

distinctive stimuli facilitates discrimination learning with stimuli that lie along the same dimension but are among themselves less distinctive. The improvement in discrimination was explained by postulating that the pretraining with distinctive stimuli aids the development of appropriate orienting responses which transfer to the more difficult discrimination. Procedurally, this approach is not unlike that used by Uznadze (1966) in his investigations of the psychology of set.

Other investigations have included the pairing of single letters with irregular inkblot shapes (Arnoult, 1953), the pairing of nonsense labels with random shapes (Ellis, 1968) and letters with 6 colours ranging from red to yellow (Cantor, 1955). By using patterns of lights in a 5 x 5 matrix, Murdock (1958) was able to investigate the role of stimulus similarity (some patterns were more similar than others) however, the results were inconclusive. In general, there appears to be fairly reliable evidence that the pairing of distinctive words, syllables or other stimuli with otherwise indistinctive stimuli facilitates the learning of new responses to the labelled stimuli. We cannot infer that the labels enhance the distinctiveness of the stimuli however; these studies merely show that the total stimulus complex stimulus-plus-label is more easily discriminated from another similar stimulus with a different label than when the same two stimuli are presented to subjects without attached labels. This finding in itself is interesting and suggests that the role of labelling in the perception of stimulus relations may be of interest in a detailed

investigation.

### 3. Studies of Verbal Pretraining in the Acquisition of Motor Skills.

Closely related to the predifferentiation studies are investigations in which motor acquisition is considered to be facilitated by verbal factors which do not exert their influence at the discrimination stage of learning, but rather, facilitate learning of the correct responses to the already-discriminated physical stimuli. In "stimulus predifferentiation" studies subjects learn to associate words or other symbols with stimuli which are later used in the motor task. In the "verbal pretraining" studies subjects associate words or phrases with stimuli which substitute for those later used in the motor task - the replacement words or phrases in fact indicate to subjects the correct motor response to be made to each stimulus. McAllister (1953) has clarified the differences between the two approaches and suggested the following new terminology:

- (i) Irrelevant pretraining, in which the training makes use of irrelevant stimuli or response words or both,
- (ii) Relevant S, in which pretraining involves relevant stimuli but irrelevant response words,
- (iii) Relevant S - R in which pretraining involves both relevant stimuli and relevant response words. Relevant stimuli are the same as, or are reasonable substitutes for motor task stimuli while relevant response words are those which indicate, for the

motor task, the appropriate responses to particular stimuli.

Relevant S as defined above is "stimulus predifferentiation" while relevant S - R is verbal pretraining.

McAllister tested the influence of three kinds of relevant S - R training in which subjects had to learn to associate one of 6 colours with each channel of a star discriminator. Groups of subjects associated the colour patches with response words based on one of three analogues of the star channels on the discriminator: there was a "degrees" analogy (e.g. red - "300 degrees"), a "clock" analogy (e.g. red - "10 o'clock") or a "directions" (e.g. red - "left, forward") analogue. In all cases, relevant S - R pretraining facilitated motor performance. The "directions" and the "clock" analogues were superior to the "degrees" analogue. A group given relevant S pretraining associated words that were unrelated to the motor task with stimuli which simulated those used in the motor task (e.g. a red patch was associated with the word "fitful"). The relevant S subjects did significantly better than the subjects given irrelevant pretraining but showed less facilitation than the relevant S - R groups.

Battig (1956) in a similar study confirmed McAllister's results. However, by using more complicated stimulus displays, Battig was able to manipulate the complexity of the task. He found a significant decrease in transfer as task complexity increased. He suggested two reasons for this:

(a) Verbal pretraining may primarily influence learning by channelling subjects' responses in a direction which leads to a

more rapid discovery of the correct solution, particularly where the simpler learning tasks are involved.

(b) Performance presumably becomes more important as task-complexity increases. There is a limit to the number of cues to which a subject can react in a complex task, so as complexity increases, verbal cues begin to interfere and verbal pretraining becomes ineffective.

Using 4 light intensities, each of which could be turned out by the subject manipulating the correct one of 4 switches, Holton and Goss (1956) investigated positive transfer to a discriminative motor task as a function of 5 types of premotor verbal experience:

- (a) The learning of different nonsense syllable names for the light stimuli,
- (b) The learning of experimenter-supplied familiar names for the stimuli (e.g. dull, very bright),
- (c) The learning of self-supplied familiar names for the stimuli,
- (d) Instructions to see, discriminate among and name the stimuli covertly, and
- (e) Instructions to see, discriminate among and name the stimuli overtly.

There were 15 groups of 10 subjects including a control group given no premotor experience. In terms of both errors and errorless trials, the motor task performances of the groups given premotor experience were superior to the performances of the control group.

McCormack (1958) has summarised the general explanation for

the facilitation of motor learning by verbal pretraining as follows: He assumes that verbal responses generate cues ( $s_1, s_2, s_3, \dots s_n$ ) which accompany the stimuli ( $S_1, S_2, S_3, \dots S_n$ ) during the motor task. Hence when  $S_1$  is encountered, the total complex is  $S_1 + s_1$ . If  $S_1, S_2$  and  $\dots S_n$  are highly similar, and if  $s_1, s_2$  and  $\dots s_n$  are distinctive, then discrimination will be easier between  $S_1 + s_1$  and  $S_2 + s_2$  than between  $S_1$  and  $S_2$  and motor performance will be facilitated. He reports an experiment by Rossman and Goss (1951) in which small amounts of pretraining retarded rather than facilitated motor learning. He assumes that, with small amounts of pretraining, each stimulus elicits each label equally often so that discriminative motor responses are made between  $S_1 + s_2$  and  $S_1 + s_1$  and between  $S_2 + s_1$  and  $S_2 + s_2$ . This is the same in effect as adding an identical component to two already similar stimuli; hence discrimination is delayed and motor performance is retarded. Using 6 similar red-yellow hues paired with star discriminator channels he showed that subjects pretrained on the motor task stimuli gave inferior performances early in the task but superior performance later.

Duncan (1953) using a slightly different approach studied transfer using a star discriminator paired with 6 lights. Transfer was studied as a function of 2 variables; degree of learning of the first task and similarity of the 2 tasks. Similarity between tasks was defined as the number of lights (2, 4 or 6) which were newly paired with different discriminator channels on the final task. Transfer increased both with degree of first task learning and



inter-task similarity. Transfer was attributed to a combination of response generalization and "learning to learn".

The studies of verbal pre-training lead us to much the same conclusion as we arrived at in the review of the stimulus pre-differentiation studies. That is, the labels facilitate learning but only by providing an additional stimulus cue to which a learned response can be made. We do not need to postulate a change in the perceived relationships between the unlabelled and the labelled stimuli. In particular there is no need to postulate that the stimuli themselves are made more distinctive by having labels paired with them; this problem is related to the perception of stimulus relations and will now be discussed in some detail.

#### 4. Studies of Transposition and the Perception of Stimulus Relations.

Reese (1968) refers to transposition as a kind of transfer that appears to result from response to relations among stimuli or to patterns of stimulus qualities rather than to the absolute qualities of the stimuli, for example the continuing choice of the larger of two stimulus objects after changes in their absolute sizes. The term mediation appears to be invoked frequently in discussions on the perception of stimulus relations; likewise the role of similarity between stimuli is an important consideration in discussions on stimulus relations.

Reese (1968) summarises a number of studies and comes to the following conclusions: Increasing stimulus similarity generally

increases transposition in both human and animal subjects. However, the effect appears to reverse for young "verbal" children; they show more transposition with distinctive stimuli than with very similar stimuli. This inconsistency may be due not to a reduction in transposition with similar stimuli, but to an enhancement of transposition with distinctive stimuli. The young child may be able to utilise some kind of verbal concept to mediate the transposition response only when the stimulus differences are large, perhaps because small stimulus differences fail to arouse the conceptual response. In a test of the relationship between stimulus similarity and transposition, McKee and Riley (1962) found more transposition of loudness than of pitch. A series of related studies, reported in Reese (1968) led these authors to conclude that although discrimination between different pitches was found to be much more difficult than between different loudnesses, the observation that more transposition occurs between the stimuli which are harder to discriminate could be attributed to verbal factors. The explanation was that children may have more verbal labels for different loudnesses than for different pitches, and that these labels could in some way mediate transposition, perhaps by inducing an "acquired equivalence" of cues, during the training and test phases.

The mediation theory interpretation of transposition has been discussed in some detail by Reese (1968). Theoretically, if stimuli arouse verbal responses or labels in subjects, then approach and inhibitory tendencies can become conditioned to these labels.

Suppose we train a subject to approach a 2" square and avoid a 1" square. The 2" square arouses the label mediating "larger" which then becomes conditioned to the approach response, while the 1" square arouses the mediator "smaller" which becomes conditioned to the avoidance response. Now on the test trial we present an eight inch and a four inch square. By generalisation the 8" square arouses the mediator "larger" while the 4" square arouses the mediating response "smaller" and so the conditioned subject approaches the former square and avoids the latter. The theory has aroused a certain amount of criticism, but it is difficult to test experimentally. Nevertheless, the concept of mediation has been considered to provide a useful description and explanation of much of the data from transposition studies.

##### 5. Mediation in Concept Formation Studies.

We next turn to what may be the most thoroughly investigated field in which mediation is considered to play a part - the role of mediating stimuli and responses in concept learning. The early research in this field has been thoroughly summarised by Goss (1961). On the basis of the studies discussed, plus his own research, he concludes that concept learning, as typically demonstrated in the laboratory, cannot be conceived without the postulation of common verbal or other responses to subsets of stimuli whose members are highly dissimilar physically (but not conceptually). Fenn and Goss (1957) used sets of blocks which varied on a number of dimensions (e.g. height, thickness, colour)

and paired exemplars of certain of the dimensions with names which were either real (e.g. small, but tall) or nonsense syllable (e.g. CEV) names. Their prediction, using blocks which had been labelled in terms of height and size, was that "The common mediating stimuli should enhance similarity of the blocks within each category. Acquired distinctiveness based on different mediating stimuli for each category should decrease inter-category generalization based on colour and shape. Therefore, these relationships of external and mediating stimuli should increase the probability of manipulative sorting of the blocks into height-size categories".

The results showed that experience with familiar word and nonsense syllable labels led to a significant increase in sortings of the blocks on the basis of height and size. Both nonsense and real word labels were equally effective as mediating stimuli. The experiment was repeated using pre-school children (Carey and Goss, 1957) with the result that familiar word but not nonsense label mediating stimuli produced the results predicted. A third investigation (Lacey and Goss, 1959) investigated variations in the mediated similarity of the blocks in the 4 height / size categories. For separate groups, 4, 3 and 2 nonsense labels were assigned to the blocks. It was expected that the 4-label condition would produce the greatest dissimilarity among the 4 height-size categories, and as predicted, sorting was best performed by those subjects for whom the labels had emphasised the dissimilarity between the 4 height / size categories.

Namikas (1967) investigated the effects of relevant, neutral,

and irrelevant pretraining with respect to the dimensions involved in his concept formation task. The results showed that relevant pretraining facilitates concept identification, and also that informational feedback about the correctness of each response improves performance.

Johnson (1967) in an attempt to uncover the nature of the mediational responses in concept identification explored the differences between reversal shifts, intradimensional shifts and extradimensional shifts. An extradimensional shift involves a change in both the relevant dimension and the values on the dimension (e.g. from red and orange colours to triangular and square shapes.) An intradimensional shift involves only a change in the levels of the same relevant dimension (e.g. from red - yellow to blue - green colours). A reversal shift involves a change in neither. If the subject acquires, during initial training, a mediation response which corresponds in some manner with the relevant dimension, the mediation response would remain appropriate after a reversal shift, but in the case of an extra-dimensional shift the response would have to be extinguished and a new one learned. We can also ask whether the mediation response corresponds more closely to the relevant dimension (e.g. colour) or to the values of stimuli on the relevant dimension (e.g. red - green). Unlike an extradimensional shift, an intradimensional shift involves a change only in the levels of the same relevant dimension, hence if mediation responses are specific to the particular levels of the relevant dimension, then an intradimensional shift would approach

an extradimensional shift in difficulty. The data analysis showed that reversal and intradimensional shifts did not differ significantly in trials to criterion, however the extradimensional shift required significantly fewer trials. This suggests that mediational learning which corresponds to the relevant dimension is important in concept learning.

In general, then, there appears to be good evidence that verbal factors play a part in conceptual learning. Kendler (1961) has defined concept learning as any situation that involves the "acquisition or utilization or both of a common identifying response to dissimilar stimuli". Hunt (1962) defines concept learning as the "learning of names", where a name is a symbol used to refer to a set. In fact, one of the characteristics of language is that words generally refer to classes of objects rather than to discrete examples or instances; it is not surprising therefore, that language should be a relevant factor in conceptual learning. Granted then, that verbal factors play a part in concept learning, the question to ask is whether verbal factors are actually a necessary part of cognitive activity. This problem will be discussed further at a later point.

## 6. Studies of Mediated Generalization.

Relevant literature in this field has been summarised up to 1960 by Horton and Kjeldergaard (1961). Problems of mediated association have been discussed elsewhere. A special application of the theory of mediated association however, can be invoked to

explain why two different stimuli elicit the same response (stimulus equivalence), why two different responses are evoked by the same stimulus (response equivalence) or why a stimulus and response elicit each other without having been previously directly associated (chaining). Most of the proposed mediational models now current have been derived from Hull's notion of secondary generalization. Horton and Kjeldergaard (p.2) classify them into two distinct classes:

- (a) representational mediation models such as those of Cofer and Foley (1942), Osgood (1952), and Mowrer (1954), and
- (b) associational mediation models like those of Jenkins (1955) and Russell (1955). The former variety emphasises the nature of the mediation process, while the latter kind emphasises its effect.

Although the history of mediate association can be traced back to Hull's work in the 1930's and the work of Shipley (1933, 1935) and Lumsdaine (1939), Horton and Kjeldergaard (p.10) report that a surprisingly small amount of research has used paired associate learning of verbal material of the type discussed earlier. Such work suggests that factors such as reverse associations and contiguity are important in mediation.

Eight paradigms of mediated generalization are developed and investigated experimentally. Analysis of the possible mediating links in these paradigms led the authors to the conclusion that backward or bidirectional association was a necessary explanatory construct to account for mediation. They suggested that task set (e.g. what is the subject supposed to be doing) and strategy (how

the subject goes about doing it) are possibly important variables in determining the significance of mediation. (p.24.)

Koplin (1967) points out that in most of the earlier studies generalisation was tested between associated words which were selected from free association norms. He attempted to produce mediated generalization of a lever-pressing motor response by means of experimentally induced or acquired associations, however was unable to do so. This failure was attributed to a "pool effect" in which a cohesive pool of bonds is established between all stimuli used during the training for acquisition of the motor response. Malloy and Ellis (1970) established a link between the predifferentiation studies and the mediated generalization approach by showing that training designed to produce acquired distinctiveness or acquired equivalence of cues produced an increase in the tendency for subjects to give generalised responses to test stimuli. This finding is consistent with the view that attaching common verbal responses to different stimuli generates common response-produced cues which increase the functional equivalence of the stimuli.

This, then, is a survey of the main fields in which mediation is postulated to play a part. An attempt will now be made to draw some general conclusions about the nature of mediation so that the theoretical background to the research described in later chapters can be sketched. Broadly speaking, in all the mediation studies reported here, the subject faces a specific task, usually involving learning. The mediating stimuli and responses appear to



facilitate acquisition of the correct response to the task by providing additional relevant cues to the subject which either channel his behaviour in the direction of a new strategy or else facilitate learning by providing additional relevant associations which serve as cues to direct his behaviour.

One point which emerges from the studies summarised here is that there is little agreement between theorists on the process to which the term mediation should refer. Bugelski (1962) uses the term to refer to a process where subjects actively establish mnemonic words or phrases in paired associate learning, e.g. the word "DEPUTIZE" to mediate the pair "DEP" - "TEZ". On the other hand, Osgood, Suci and Tannenbaum (1957) suggest that a representational mediational process may be an element of all S - R learning. They divide the usual S - R paradigm into two stages - decoding (the association of the stimulus with representational mediators) and encoding (the association of the mediated self-stimulation with overt instrumental responses). Staats (1968, p.188) takes the view that language mediates behaviour because language and thought are synonomous:

"to say that language sequences . . . . determine thought is redundant. These language sequences are types of thought and could be called chains of thought responses. It is suggested that a large portion of what we call thought is composed of such chains".

Staats, along with Osgood (1953) and Mowrer (1954) views word meaning as an implicit response. Osgood suggests that meaning

responses come under the control of the word stimuli through instrumental conditioning, while Mowrer and Staats suggest that the principle involved is that of classical conditioning. For these theorists then, language mediates thought because thought is largely the manipulation of words to which implicit meaning responses have become conditioned. This view of the role of verbal processes in thought has received some experimental support. For example, Staats and Staats (1957) took the view that if meaning is an actual response, it should be possible to classically condition it, i.e. the meaning response elicited by a word could be conditioned to any contiguously presented stimulus. In three experiments a nonsense syllable was presented 18 times, each time paired with the auditory presentation of a different word. While these words were different, they all had an identical meaning component (as measured by the Osgood semantic differential scale). In each experiment there was significant evidence that meaning responses had been conditioned to the nonsense syllables. (Data from those subjects who reported "awareness" of a relationship between certain words and syllables were not analysed.) Brown (1958, p.101) has criticised the conditioning approach:

"Osgood . . . pushed the mediating response back into the nervous system where no one will undertake to look for it. . . . The theory is not to be judged on the evidence for fractional implicit responses but by the success with which it predicts . . . overt behaviour. I have found that this success cannot, at present, be

given anything approaching a conclusive evaluation. Conditioning theories of linguistic meaning have been squeezed by the same pressure to the same sanctuary - the central nervous system. Assuming that words must have immediate effects, and plagued by a shortage of overt effects, the conditioning theorist has revised his conception of meaning from overt response to implicit and eventually mediated response."

A similar criticism of the conditioning approach has been put forward by Wallach (1958). The verbal conditioning approach also came under fire in Chomsky's (1967) critique of Skinner (1957). In addition, recent experimental studies, e.g. Maltzman, Langdon and Feeney (1970), Maltzman (1968), Feather (1965), Maltzman and Belloni (1964), Deno and Jenkins (1967) and Turquin (1969) have cast doubt on the view that conditioning via mediators can account for the "meaning" of verbal stimuli and responses.

There appear to be few adequate definitions of mediation in the literature. The most satisfactory definition noted by the present writer is that given by Duncan (1967, p.347):

"Mediation involves the apprehension of an association or connection of some kind between words or other events."

In fact, it seems reasonable to reduce the whole of the mediation concept to the postulate that mediation involves the awareness of some additional piece of learned information which may or may not be explicitly or overtly given in the experimental situation but which serves as a cue for the subject in his task of solving

the experimental problem set for him by the experimenter.

If we accept the view stated by Garner (1962), that uncertainty is a prerequisite to structure and that the "search for structure" is inherent in behaviour then the additional cues provided by so-called mediating stimuli and responses can be considered to facilitate the subject's task simply by providing additional information. Gibson (1966) has taken a similar view, and indicates a subject who actively "hunts" or searches for meaning to make sense of what little information he can get.

Some recent research in the field of sensory interaction (e.g. Wilson and Gregson, 1967; Mitchell and Gregson, 1968) suggests that increasing the confusion or ambiguity in the subject's task leads to a greater reliance on any irrelevant stimulus cues that are present in the experimental situation. It could be argued that if we add labels or names to stimuli which are being handled by the subject in a relatively confusing experimental situation the labels will be used as a cue to reduce uncertainty and hence aid the subject in his search for structure.

If we accept this view of the central nervous system as an active organiser of information into structural categories, we can foresee two possible relationships between this structuring process and the structure of language itself. Since words generally refer to classes of stimuli we can postulate that either we classify the environment into groups of objects because we have verbal concepts of these classifications (i.e. language in a sense mediates perception) or we can postulate that we classify stimuli

and only then label them (i.e. we perceive first, then label). Voss (1969) has pointed out that all perceptual behaviour is inseparably related to the thought processes and vice versa. The most acceptable current view of perception appears to be the information processing viewpoint. The major assumptions of this view have been outlined by Voss (1969; Chapter 1) and include the following points:

A perceptual response is assumed not to be an immediate consequence of stimulation but rather one which has gone through a number of stages or processes, each of which takes time to organize or transverse. It is assumed that total time from stimulus onset to the occurrence of a perceptual response can be divided into intervals, with each interval characterised by a different operation. We can then create a block design of these intervals. A number of such models have been constructed, e.g. Broadbent (1958), Sperling (1963) and Melton (1963).

A second main assumption of the information processing viewpoint is concerned with limited information handling capacities. We thus need to look for instances in which recoding of information takes place in such a way that some of the content might be expected to be maintained more explicitly at the expense of other aspects which are lost. Verbal information which accompanies the stimulus input, for example, may be handled as a significant part of the total input in some circumstances, but rejected as contributing nothing to the search for structure under other circumstances.

Earlier reference was made to a trend in recent research which suggests that mediation would be better viewed as one of perhaps several alternative strategies which are available to subjects in a typical learning task. Indeed, the term mediation itself, as currently used in the literature, appears to be broad enough to encompass a wide range of possible strategies. This suggestion is given support by the observation that instructions to mediate frequently improve performance on learning tasks in which mediation is believed to occur. (Paivio and Yuille, 1967; Yarmey and Csapo, 1968; Yuille and Paivio, 1968; and Schwartz, 1969). Pelton (1969) has shown that the construction of mediation by the subject himself is more effective than the perception of a ready-made link in the facilitation of learning by mediation, again suggesting that mediation operates as a strategy.

The approach that will be taken in this thesis is that the concept of mediation has often been used as a blanket term to describe a theoretical situation which could be more adequately and more precisely explained in alternative terms. It is held that a better understanding of the process could be achieved by a closer examination of the way in which information is utilised by subjects in a learning or judgement task. Such an approach would allow us to draw upon many more diverse but nevertheless related areas of psychology which have hitherto been regarded as fields which are distinct, in both methodology and in terms of underlying processes, from the mediation studies which have been discussed on earlier pages.

The area which is directly relevant to the problem of the role of verbal processes in judgement has been currently classified as a branch of psycholinguistics. In particular, we are interested in studies which have been concerned with the role of language in cognitive processes. It is to these studies that we now turn.

## PART II.

### THE ROLE OF LANGUAGE IN COGNITIVE PROCESSES

Subjectively, it is difficult to conceive of the possibility of thought without words; yet the psychological interpretation of the role of language in thought remains an area of great controversy. We have seen that the current theoretical trend is against theories which connect verbal processes to cognitive activities via conditioning. The studies reviewed in this section have been less concerned with the mechanisms by which language might be considered to mediate thought; rather they stem from the area of linguistics and particularly the linguistic relativity theory of Benjamin Lee Whorf.

The two extreme viewpoints on the role of language in cognitive behaviour which were noted in the review of the mediation studies also occur in this field. The poet Mandelstam showed an interest in the problem when he wrote "I wanted to utter a word, but that word I cannot remember; and the bodiless thought will now return to the place of shadows".

McGuigan (1966, p.3.) appears to equate thought with language processes when he defines thinking quite simply as "the composite of bodily activity (neural, muscular, etc.) that necessarily results from the presentation of a language stimulus". Bruner, Goodnow and Austin (1956, p.60.) on the other hand claim that "many experiments in concept attainment, including our own, have shown that subjects are able to distinguish correctly exemplars from nonexemplars of a concept before being able to name the defining



features on which their judgements are based".

Carroll (1964) lists one of the major functions of language as a system of responses that facilitates thinking and action for the individual. He also concedes that many kinds of thinking are possible without language.

An excellent review of studies relevant to the problem of determining the role of language in cognitive processes can be found in Lenneberg (1967, p.346-363).

He summarises a number of important studies to support his view that naming is actually the consequence rather than the cause of categorisation processes. He also raises the important question of whether the semantic structure of a subject's language influences his cognitive structuring of the sensory input arising from his environment. This is the Whorf Hypothesis of linguistic relativity, which has aroused considerable interest from psychologists and linguists alike.

Carroll (1964) takes the view that possession of names for objects does not alter our absolute capacity to discriminate among these things when they are extremely similar. He adds however that the special names which some specialists learn to use to refer to stimuli such as hues help in one way; they enhance the ability of the user to recognise and identify particular instances from memory. He quotes as evidence the well-known Brown and Lenneberg (1954) study in which colour codability was related to how well a given colour could be recognised in a task where it had to be picked out from an array of 120 colours. This approach has been

discussed in detail in Lenneberg (1967).

Their measure of codability is essentially a measure of how well people agree in giving a name to a stimulus; in this case, a colour. For example high codability would be expressed by a word which uniquely and unambiguously labels a highly specific stimulus, for example the physical colour of blood. The subjects, who were all native born American English speakers, were screened for colourblindness. Their task involved the recognition of certain colours after they had been shown them one, (or in some cases, four) at a time for short periods. After a timed interval they viewed a large colour chart from which they were required to choose the colours which they had been shown before. The measure of codability, which had previously been determined, and has since been independently confirmed by Beare (1963), was not found to predict the recognisability of a colour when the recognition task was easy. (One colour to be recognised after a seven second waiting period). However, as the task was made more difficult, codability began to correlate with recognisability and the relationship was most clearly seen in the most difficult of the tasks (four colours had to be recognised after a three minute waiting period during which subjects were given irrelevant tasks to perform.) At first this looked like experimental proof that under certain conditions a person's native language may facilitate or handicap a memory function. However, it later became apparent that the way in which the colour stimuli were selected was a crucial factor. Burnham and Clark (1955) used the same

procedure as Brown and Lenneberg, but used different colours, and instead of a positive correlation between codability and recognisability they found a negative correlation. Lenneberg (1967) reports that both experiments were repeated by Lantz (unpublished) and Lantz and Stefflre (1964), all with similar results. He concludes that semantic structure influences recognition only under certain experimental conditions namely when the task is difficult and the stimuli are chosen in a certain way. In other words both of the experimental techniques used are special cases.

The Lantz and Stefflre (1964) study showed that it is not the semantic characteristics of the language that influenced the cognitive process of recognition, but rather it is the particular use that subjects will make of language in a specific situation. Instead of using the Brown and Lenneberg approach, which primarily brings out language peculiarities, they measured accuracy of communication between subjects. They showed that the accuracy with which each stimulus in an array can be communicated between subjects predicts recognition of that colour. Since codability predicts recognisability only in special contexts, Lenneberg (1967, p.354) infers that subjects make use of the ready-made reference facilities offered them through their vocabulary only under certain circumstances: "...he is in fact not bound by the semantics of his natural language; there is little evidence of the tyrannical grip of words on cognition."

Lenneberg (1967, p.355) concludes that whatever slight

limitations the semantic structure of a language may have on recognition under special circumstances, these limitations may be largely overcome by the creative use of descriptive words by the individual. He accepts the Lantz and Stefflre suggestion that there are situations in which the individual communicates with himself over time, when faced with a complex cognitive task.

Carroll (1964) reports two cross-cultural studies which were designed to test the Whorf hypothesis. In 1956, Lenneberg and Roberts replicated the Brown and Lenneberg experiment, using speakers of the Zuni language (used by Indians of New Mexico). Striking differences between English and Zuni speakers were reported both in the codability of colours and the ease with which various colours could be remembered by the speakers of the two languages. Casagrande worked with Navaho children whose language requires that certain verbs of handling (pick up, hold, drop, etc) take special forms depending on the object being handled. He compared Navaho and English speaking children matched for age, with respect to how often they used shape, form or material as a basis for sorting objects rather than colour; he used sorting tasks that are usually performed by young children on the basis of colour. The Navaho-speaking children had a tendency to perform the task on the basis of form at distinctively younger ages than the English-speaking children. However, a later study using Boston children suggested that environmental experience rather than language was the crucial variable. Carroll concludes that there is little convincing support for the linguistic relativity

hypothesis.

Other experiments have attempted to demonstrate the advantage of words in various tasks in which perceived impressions have to be stored and remembered in some way. Bartlett (1932) analysed data on retention in terms of whether or not subjects had spontaneously verbalised during observation of stimulus materials, and found that those who verbalised had superior retention. Kurtz and Hovland (1953) had children observe an array of 16 familiar objects. Half of the subjects were given a sheet of names of the objects while the remainder were given a sheet of pictures of the objects. As the experimenter indicated each object the subjects found and encircled the name or picture of the object. In the group with printed names, the subject pronounced the name aloud while encircling the object. No indication was given that there would be a later test of retention. Recall was tested after one week and superior retention was found in the verbalization group. The study can be criticised however by postulating that verbalization may require subjects to direct their attention to the more distinctive properties of the objects while the control group had only to discriminate at a level sufficient to differentiate the objects from each other but not from the new items presented in the recognition test.

A classic investigation of the role of labelling in retention was performed by Carmichael, Hogan and Walter (1932). They showed that verbal labels which are presented prior to exposures of certain ambiguous line drawings result in the reproduction of figures which

resemble the objects referred to by the labels. An early study by Hanawalt and Demarest (1939) showed that if the labels are presented immediately after or even up to seven days after presentation of the figures, the labels will still influence reproductions at each recall interval.

Bruner, Busiek and Minturn (1952) exposed the figures for 10, 50 and 100 milliseconds, with the verbal label given prior to the exposures and with subjects reproducing the figures immediately after each exposure. Again the verbal labels were found to influence figure reproduction, and Bruner et al interpreted this as evidence for perceptual assimilation as distinct from the mnemonic assimilation as found by Carmichael et al.

Prentice (1954) used a recognition rather than a reproduction technique. His subjects observed a series of verbally labelled figures twice and then were required to select the previously seen figures from a group of 60 figures which contained the 12 original figures plus 48 figures which were distorted to represent the labels of the two different word lists used. Prentice found that errors in the subjects' selections did not significantly favour the distortions that were consistent with the labels used. He concluded that the label influence is a function of the process of reproducing the figures, rather than of the subjects' perceptual interactions with them or their remembering of them.

Herman, Lawless and Marshall (1957) investigated the role of set and exposure time as variables influencing the frequency of language - influenced reproductions. Subjects who were given instructions to view the figures with the intention of

reproducing them later, drew significantly fewer language-influenced reproductions than subjects who were not given these instructions, i.e. the former group presumably paid more attention to the details in the figures. A strong tendency was also found for subjects to draw more language-influenced reproductions when the figures were exposed for shorter intervals. It was inferred that the verbal labels serve to channel the stimulus function of the figure and this stimulus function tends to be implicitly operative during reproduction of the figures. Carroll (1964) notes that the use of a label, whether by the subject alone, or also by the experimenter, implies that the label refers to a concept; thus the figure is perceived as being one of a class of similar experiences named by the concept.

Carroll (1964) takes the view that the existence of a word which names a concept in a subject's vocabulary is responsible for making a given concept particularly salient and easy to attain. Lenneberg (1967, p.356) however does not accept the postulate that natural language is a biasing factor in the formation of concepts in general. He points out that much human activity is based on concepts that must have taken place in the absence of naturally occurring words; for example, in mathematics, a language is simply created ad hoc as the concepts are developed. He sees naming as a creative process, not a rigid convention.

He comes to the conclusion (p.362) that "the semantic structure of a language may influence cognitive structuration where our physiological equipment allows for a range of alternative

solutions. But there is no indication that the basic organizational capacities are crippled through a crippling of language proficiency."

The studies of the role of language in cognitive processes, though traditionally distinct from the typical mediation experiments, may in fact differ only procedurally from the latter studies. It was suggested earlier that we abandon the old term "mediation" in favour of a new system in which we consider the subject's task in terms of additional cues and associations provided by the information formerly referred to as the mediating stimuli and responses. We can then conceptualise the subject's task as one of integrating and processing the available information in order to arrive at a solution to the problem that has been set by the experimenter. Our interest in the problem of verbal mediation in judgemental processes then becomes centred on determining the influence of verbal information, as provided by stimulus labels, on the outcome of the judged similarity between the stimuli.

Few studies appear to have investigated this problem in precisely the manner outlined above. A study by Nurminen (1965) however was conducted on similar lines to the approach proposed here. She associated 8 different Finnish names, which fell into two distinct classes, with 8 nonsense shapes, by paired associate learning. The similarity of the shapes was measured both before and after they had been labelled. There was a significant trend toward enhanced judged similarity between labelled shapes belonging



to the two classes of names, but a nonsignificant trend towards greater dissimilarity in the predicted direction, i.e. the labels enhanced similarity between shapes assigned to the same class by names but did not decrease similarity between shapes in different classes. The subjects were school children, and similarity was assessed by the direct approach. Other studies have used indirect techniques to investigate the effect of labels on similarity, e.g. Katz (1963). The results, however, tend to be confounded by the facts that similarity is usually measured simply as a bipolar judgement of "same" or "different" or the dependent variable is either the acquisition of a discriminatory response (often motor) or a measure of stimulus generalization. Katz and Zigler (1969) have also suggested that age of the subject is a relevant variable.

There is some evidence then, that similarity judgements can be manipulated by appropriate pairings of labels with unfamiliar and hence potentially un-named stimuli. This raises a number of questions of theoretical interest. For example, do nonsense labels produce greater modification with unfamiliar, as distinct from familiar stimuli? What is the magnitude of shift in judged stimulus similarity when labels of various kinds are paired with stimuli? Problems concerned with the underlying dimensionality of the stimuli also arise when we add labels to stimuli. If it can be shown that labels increase the dimensionality of the similarity judgements, then we need to ask whether this increase is due to the labels as such or simply to the fact that we have added information to the stimuli, in much the same way as if we

had added a dimension of colour to taste stimuli. Other questions of interest are concerned with the way the stimulus plus the label are combined to produce an overall judged similarity. This thesis is concerned with an investigation of problems such as these. The theoretical background and an experimental investigation of some of these points will be examined in the chapters which follow, and the relevance of the results to our understanding of the relationship between cognition, labelling and judgement will be discussed where appropriate.

## CHAPTER II

### AN INITIAL INVESTIGATION OF THE EFFECTS OF NONSENSE LABELS ON TASTE SIMILARITY JUDGEMENTS.

The previous chapter included a review of the research that is directly relevant to the problem of determining the role of verbal processes in complex cognitive behaviour. One feature which emerged from the studies reported is that there appear to have been few attempts to investigate the effects of linguistic manipulation on individual subjective judgements.

When we attempt to investigate the role of verbal factors in such processes as discrimination and judgement, a number of problems arise. In very general terms, the theoretical situation which is of interest to us is one in which we have a finite set of stimuli, which can be perceived to vary along a number of distinct dimensions or continua. We instruct the subject to structure these discrete stimuli as a series of points along one such continuum which he then does according to some intrasubjectively determined rule. For example, he may be required to assess the relative similarity of the stimuli, or he may be asked to rate them in terms of hedonistic appeal. The verbal manipulation that we wish to investigate consists of an attempt to alter the outcome of this rule by pairing some or all of the stimuli with a set of names or labels which differ from the labels that the subject is likely to have paired with the stimuli in the absence of the experimental manipulation. The experimenter

does not attempt to alter the rule by which the stimuli are ordered by a given subject (though he may in fact alter the way in which the rule is applied). Rather, he attempts only to alter the ordering of stimuli within the continuum by actively manipulating the labels assigned to the stimuli. We can, of course, expect to produce shifts in relative orderings of the stimuli only if the stimulus labels carry information that is relevant to the subject in his task of ordering the stimuli, and this additional information causes the subject to revise his ordering of the stimuli.

An important factor which relates to the general experimental technique outlined above is the problem which will be referred to as "overwriting". Given that we can overtly change the labels paired with stimuli, we assume that if "overwriting" occurs, then we have in some way altered the subject's internal verbal categorisation of the stimuli with which the labels are paired. This altered categorisation then becomes the basis for a new subjective ordering of the stimuli within the continuum that we wish to investigate.

To be more specific, suppose we wish to manipulate the labels that are paired with a set of taste stimuli. We have decided that the subject's cognitive structuring or ordering of the stimuli which we wish to investigate is to be in terms of similarity of the stimuli to each other. That is, we use the subject's judgements of relative similarity as our dependent variable, and predict that our verbal manipulation of the stimulus labels should produce measurable shifts in the similarity judgements generated by the subject. Initially, we wish to answer two related questions:

(1) Can we demonstrate that "overwriting" of a subject's own previously-learned labels or categorisations of a set of stimuli occurs when new labels are experimentally paired with the stimuli?

(2) Given that "overwriting" occurs at the subjective level, can we empirically demonstrate that systematic shifts in the ordering of stimuli within a subject's continuum of perceived similarity are produced by manipulating the labels paired with individual stimuli?

It should be noted that the effect which we are investigating is necessarily second order.

As Staats (1968) has pointed out, both the language used by an individual, and his experience of the world in which he lives, are to a large extent isomorphous with the physical environment. If then, we attempt to manipulate verbal processes independently of the physical processes which they describe, then we will produce a conflict situation in which the isomorphism breaks down and the linguist correlates of a physical event will no longer be seen as applicable or relevant to the subject's physical experience of that event. It is necessary, therefore, to study the effects of labelling under conditions in which the credibility of the situation will not break down, such as in judgement processes involving easily confused stimuli, or in situations for which the subject's sensory experience of physical events contains a lack of structure or a high degree of uncertainty. Even under such conditions we would expect, for the rational judge, that the effects of verbal manipulation will always be relatively small; perhaps something no larger than the traditional time-order effects of psychophysics. In other words, we would expect

that there is a limit to the effect that verbal stimuli can have on cognitive processes and one of our problems is to find out just what that limit actually is.

The experimental pilot study described in this chapter was designed partly to investigate the points mentioned above, and partly as a very broad and general investigation of the role of verbal labelling in the formation of subjective similarity judgements. The experiment will now be described in detail.

#### EXPERIMENT 1. (PILOT STUDY).

The following experiment was performed in the University of Canterbury Gustatory Psychophysics Laboratory. (See Gregson, 1964, for a description of the relevant features of this unit.) The subjects were 20 undergraduate students from introductory psychology courses; they were serving for partial fulfilment of their course requirements. All subjects were tested individually, and none served more than once. Subjects were seated in a small cubicle behind a pair of sliding panels which were used to close off the service hatchway; they were thus able to make their responses in isolation and out of view of the experimenter. All responses were made to a strict time schedule which was signalled to the subject via coloured lights and a buzzer. The complete cycle of events on each trial, including an encoded stimulus description, was recorded on an Esterline Angus paper chart recorder via Massey Dickinson modular units.

A set of 8 taste solutions was prepared as follows:

(A) Stimuli 1 to 4 will be referred to as simple tastes, since they contained only one substance (sucrose, sodium chloride, citric acid, or quinine sulphate) dissolved in de-ionised water (resistance greater than 2 megaohms per centimeter). An attempt was made to use concentrations that were of approximately equal subjective intensities (Beebe-Center and Waddell, 1948; Gregson, 1963). These concentrations, which are all well above threshold, were as follows:

Stimulus Number	Substance	Concentration
1.	Sucrose (Suc.)	0.12 gm/ml.
2.	Sodium chloride (NaCl)	0.03 gm/ml.
3.	Citric acid (Cit. A.)	0.005 gm/ml.
4.	Quinine sulphate (Quin.)	0.00014 gm/ml.

(B) Stimuli 5 to 8 were mixtures of the 4 simple tastes. They were prepared by mixing 3 of the above tastes, in equal volumes but diluted to half the concentrations listed above, with a 4th undiluted or "dominant" component. (The 4 components, 3 diluted and one undiluted, were mixed in carefully measured equal volumes.) This produced a set of 4 mixtures which will be referred to as:

Stimulus Number	Description
5.	Sucrose dominant mixture (Suc. mixt.)
6.	Sodium chloride dominant mixture (NaCl mixt.)
7.	Citric acid dominant mixture (Cit. A. mixt.)
8.	Quinine sulphate dominant mixture (Quin. mixt.)

The mixtures can be considered as impure or "noisy" versions

of the 4 simple tastes.

The experiment consisted of a 2 x 2 factorial design, with the 20 subjects divided equally among the 4 experimental treatment groups. In condition 1, stimuli 1 to 4, the simple tastes, served as standard stimuli against which the entire set of 8 tastes, as comparison stimuli, were judged. In condition 2, the 4 mixtures were used as standards, while the comparison stimuli were again the entire set of 8 tastes. The ten subjects in each of condition 1 and condition 2 were further subdivided into experimental and control groups. In the two experimental groups (one group in condition 1, the other in condition 2) the subjects were instructed to use a set of 4 experimenter-provided nonsense syllable names or labels to refer to the 4 standard stimuli. These labels were "zum", "tov", "jeg" and "dax". The actual one-to-one pairing of labels with tastes was varied cyclically from subject to subject. The two control groups however, were required to generate their own real-word labels for the 4 standards in each particular condition. A ten-point response scale of similarity / dissimilarity was used by the subjects. It consisted of the following categories:

Identical

Nearly Identical

Very Similar

Similar

Slightly Similar

Slightly dissimilar

Dissimilar



Very Dissimilar

Extremely Dissimilar

Completely Opposite.

These categories were paired with a set of 10 morse keys located to the right of the seated subject; a cuspidor was located on his left and ample de-ionised water was provided for the subject to rinse his mouth after each trial. The taste stimuli were presented as 10 ml. samples in 50 ml. beakers which had been raised to a constant 25 degrees Centigrade in a water bath prior to being presented to the subject. The stimuli were presented with a minimum interval of 45 seconds between each taste.

The experiment can be divided into 3 parts:

PART 1.

The stimuli which were to be used as standards were presented twice to the subject; he tasted them in the order in which they are listed above. The following written instructions accompanied Part 1 for the experimental group subjects:

"A set of four tastes will be presented to you. Please taste each in turn and make a mental note of the way in which they differ.

The same four tastes will next be given again. Please taste each in turn and I will assign a name to each so that you can verbally identify it and distinguish it from any other tastes which may be given later in this experiment. When you have learned the four names for the four tastes, write the names on the sheet of paper provided taking care to write them in the same order as you

first tasted them."

Instructions for the subjects in the control groups were identical, except that the words "I will" were omitted from sentence two. (Subjects in the control groups generated their own labels for the tastes.) Subjects listed the 4 appropriate names on a sheet of paper which remained in front of them for reference throughout the remainder of the experiment.

## PART 2.

The purpose of this section of the experiment was to familiarise subjects with the standards that they had just named, and also to provide a simple test to check that subjects could in fact distinguish each taste and pair it with its name. It also served to familiarise subjects with the key-pressing sequence. The instructions, for all subjects, were as follows:

"You will now be tested to check that you are able to distinguish each taste and pair it with its name. It will also be necessary for you to state how similar you consider each taste is, to every other taste in the series. You will make these judgements by responding with the set of keys on your right. Suppose that you named the 4 tastes:

(a)

(b)

(c)

and (d).

You will always make your comparisons and respond in the sequence (a) (b) (c) and (d), as you have just written on the paper.

Thus, if you decide, for example, that the first taste which I give to you is (c); then you may decide that it is (i) "similar" to (a); (ii) "completely opposite" to (b); (iii) of course it will be "IDENTICAL" to (c), since you have just decided that it is in fact (c); and finally you may decide that it is, for example, "very similar" to (d).

Thus after receiving and tasting this first solution you will, on the basis of these decisions make the 4 key-presses:

- (i) "similar"
- (ii) "completely opposite"
- (iii) "identical"

and (iv) "very similar".

And so on for each new taste as I present it to you. Your judgement of what the taste actually is, will always be given by pressing the key labelled "identical" but do not forget that responding is always in the sequence described above: (a), (b), (c), then (d).

After you have made the four key-presses, please tell me the name of the taste which you have just experienced. Remember you may only identify the taste by using one of the four names which you wrote down at the beginning of this experiment. When you have correctly identified each solution in this way twice, we will move on to the next part of the experiment."

This time the standards were presented 4 times each and in random order:

Condition 1				Condition 2			
(Simple standards)				(Mixture standards)			
4	1	2	3	8	5	6	7
2	3	1	4	6	7	5	8
3	4	2	1	7	8	6	5
1	4	3	2	5	8	7	6

(1 refers to sucrose, 2 refers to sodium chloride; 5 is sucrose dominant mixture, etc. as listed earlier.)

The psychophysical method employed will be referred to as the "memorised standards" method. Basically, the subject's task was to judge a given comparison stimulus (held in the mouth) against each of the previously encountered standards (fixed in the memory). The stimulus being tasted was always judged against each one of the 4 memorised standards in the order 1, 2, 3 then 4 in the case of condition 1 subjects or in the order 5, 6, 7 and 8 for subjects in condition 2. After judging each taste in turn against the 4 memorised standards, the subject was asked to name the currently experienced taste, using one of the 4 names written in front of him. (His sequence of key-presses could also be taken as an indicator of whether or not he correctly identified the taste, since on each trial, one of the memorised standards was in fact identical to the comparison stimulus that he was currently tasting; and hence one of the 4 responses on each trial should have been near the "identical" pole of the similarity scale.) The sequence of events on each trial followed a time schedule which is outlined below:

Time (seconds)	Signal	Responses
0 - 24	Red light.	Rest.
25 - 29	Yellow light.	Subject takes comparison stimulus in beaker from rack, prepares to taste.
30	Green light.	Subject takes contents of beaker into mouth and holds it there while responding.
31 - 45	4 buzzes. (Green light)	Each buzz signals subject to judge the comparison stimulus successively against the 4 memorised standards.
46 - 60	Red light.	Subject spits out solution, rinses with de-ionised water and waits for next trial to begin.

The 32 practice judgements made in this part of the experiment were not analysed; their function was merely to habituate the subject to the apparatus and the experimental technique.

### PART 3.

The instructions for Part 3 were as follows:

"A series of 32 tastes will now be presented. These may be (a), (b), (c), or (d), or mixtures of two or more of these 4 named tastes. To each one of the 32 give four responses on the scale, according to its similarity to (a), (b), (c), and (d) always considered in that order."

Each of the 8 comparison stimuli was presented 4 times, in random order, generating  $32 \times 4 = 128$  responses. The randomisation

of the comparison series for all subjects was:

7 6 5 4	1 8 2 3	6 8 5 7	2 3 1 4
4 7 3 1	5 2 8 6	1 4 6 3	2 5 7 8

The starting point in the series was varied cyclically for successive subjects. Responses were made in the time sequence listed above and the sliding panels were closed to allow subjects to respond in isolation. Subjects from condition 1 (simple standards) encountered the mixture comparison stimuli for the first time in Part 3; likewise, the condition 2 subjects first encountered the simple tastes during this final part of the experiment.

## RESULTS AND ANALYSIS.

The data were transferred from paper-chart continuous records to punched cards for analysis on the University of Canterbury I.B.M. 360/44 Computer. The comprehensive programs used in analysis of the data were prepared in their initial versions by Professor R.A.M.Gregson, to whom thanks are due.

Data from the 4 repeated presentations of each of the 8 comparison stimuli were averaged within subjects to give a 4 (standards) by 8 (comparison stimuli) matrix of mean responses for each of the 20 Ss. The variances of the 4 entries that were averaged in each cell of the 4 x 8 matrix were also calculated.

By averaging across the 8 mean judgements in each of the 4 rows of the matrix, 4 baseline scores were obtained. A second 4 x 8 matrix was calculated in which the mean responses were expressed as

z-scores about the baseline values obtained for each of the 4 standards.

Finally, for each subject, the grand mean similarity and grand variance were calculated by pooling the entire set of 128 responses produced by that subject.

The entire analysis was then repeated using data obtained by pooling the responses made by the 5 subjects in each treatment. Much of this discussion will be limited to a consideration of the pooled results. For each experimental treatment, the 4 x 8 matrix of nonstandardised similarity judgements was further condensed by combining symmetrical pairs of comparisons to give a set of 26 judgements between 26 stimulus pairs.

TABLE 2:1

Condition 1 Experimental Ss (Pooled data)

	Comparison Stimuli							
	1	2	3	4	5	6	7	8
Standard 1	1.450	7.875	8.525	9.350	5.250	7.650	7.650	7.250
Standard 2		1.600	6.425	7.400	6.000	3.250	4.950	5.400
Standard 3			2.750	5.325	6.250	5.350	4.550	5.550
Standard 4				2.000	7.250	6.850	6.150	5.550

TABLE 2:2

Condition 1 Control Ss (Pooled data)

	Comparison Stimuli							
	1	2	3	4	5	6	7	8
Standard 1	1.250	8.550	8.825	8.550	5.200	7.900	7.950	7.850
Standard 2		1.550	6.000	7.300	6.150	4.350	5.150	6.050
Standard 3			2.100	6.350	6.550	4.750	5.150	4.950
Standard 4				2.350	6.800	6.900	6.550	4.200

TABLE 2:3

Condition 2 Experimental Ss (Pooled data)

	Comparison Stimuli							
	1	2	3	4	5	6	7	8
Standard 5	2.850	6.800	7.325	8.000	3.750	5.350	5.800	5.850
Standard 6		3.150	6.525	6.675	4.650	4.500	4.950	4.950
Standard 7			5.800	5.150	5.750	4.750	4.400	4.850
Standard 8				4.750	7.000	5.450	4.950	4.350

TABLE 2:4

Condition 2 Control Ss (Pooled data)

	Comparison Stimuli							
	1	2	3	4	5	6	7	8
Standard 5	2.800	7.125	8.250	8.550	3.950	5.600	6.450	6.600
Standard 6		2.800	7.100	7.500	6.450	3.650	6.000	6.350
Standard 7			4.500	6.425	5.850	4.100	4.250	4.850
Standard 8				5.750	5.700	5.300	5.400	4.500



FIG 2:1

POOLED DATA FROM EXPERIMENT 1 - (CONDITION 1)

EFFECT OF NONSENSE LABELS ON JUDGED SIMILARITY OF TASTES.

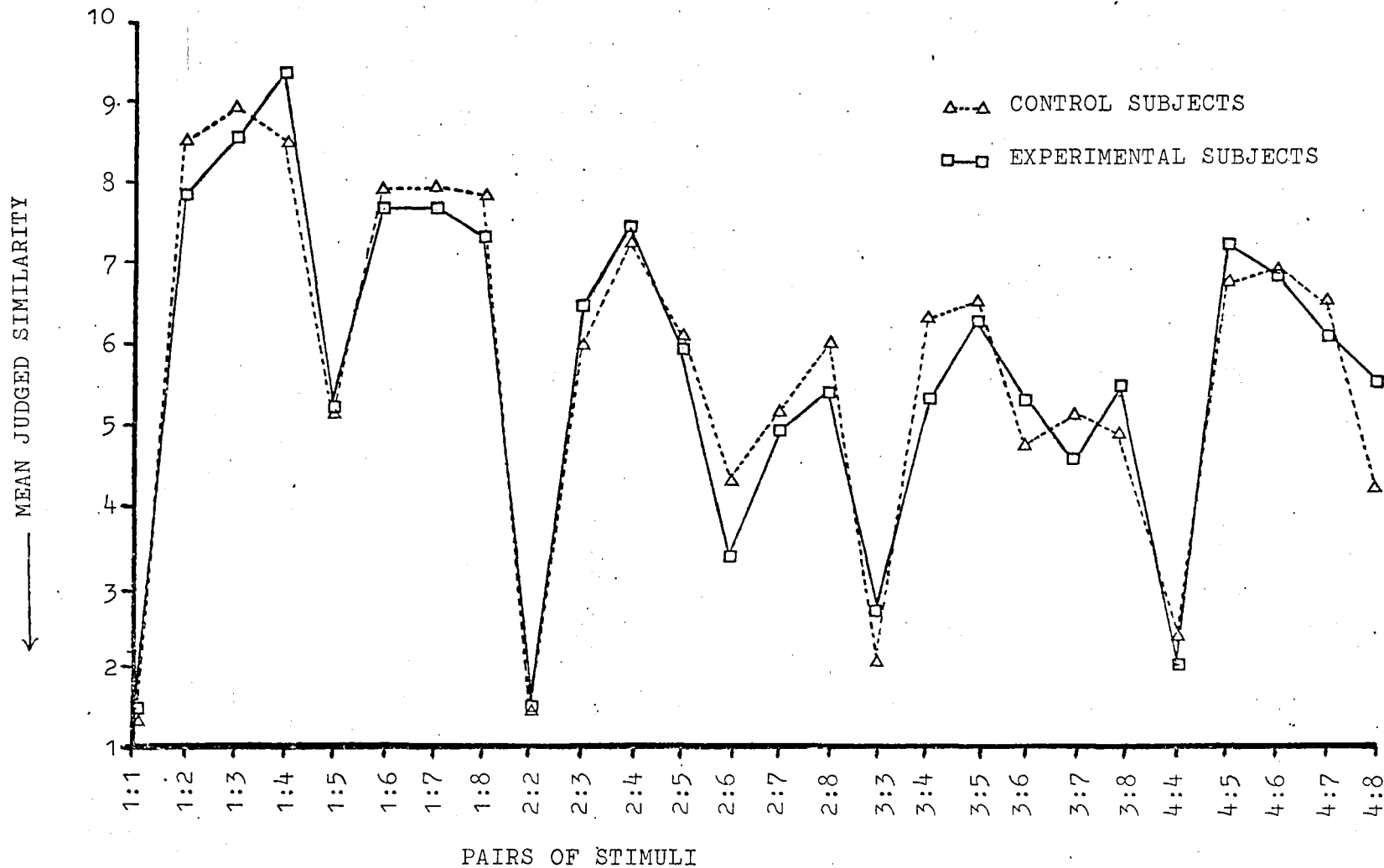
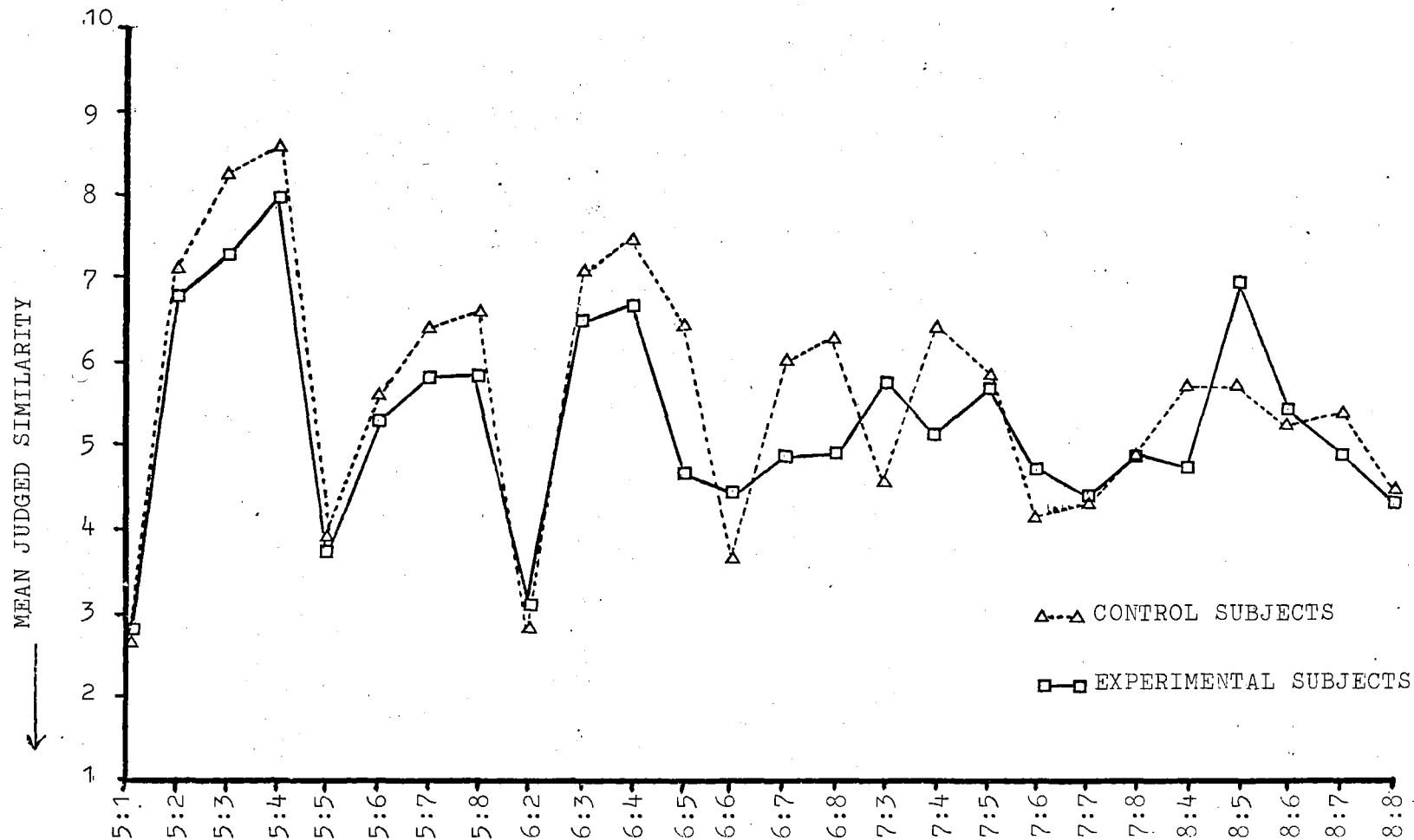


FIG. 2:2

POOLED DATA FROM EXPERIMENT 1 (CONDITION 2)

EFFECT OF NONSENSE LABELS ON JUDGED SIMILARITY OF TASTES.



The pooled data from the two treatments in condition 1 and the two treatments in condition 2 are graphed in figures 2:1 and 2:2 respectively. The same data appear in Tables 2:1, 2:2, 2:3, and 2:4. (The standardised data were also graphed, however the bunching of points around the middle of the scale and the more frequent crossing over of the experimental and control group plots made interpretation difficult. It was therefore decided to restrict this part of the discussion to the raw or nonstandardised data only.)

Both graphs show that the labelling effect is only slight when compared with the range of responses across the whole of the similarity scale. The graphs from conditions 1 and 2, which are in effect two separate but parallel experiments, bear a striking resemblance to each other in overall profile suggesting that much the same judgements are produced regardless of whether subjects use the simple tastes or the mixtures as standards. There is, however, a noticeable flattening of the graph in condition 2; this can be attributed to the fact that the mixture standards had more physical attributes in common and also were more difficult to discriminate and hence were judged to be more similar to each other. Furthermore, the error variances in the repeated measures were considerably higher in condition 2, suggesting a more frequent misidentification of the taste stimuli in that condition.

The factor of greatest interest here, however, is the magnitude of the relative shifts in judgements between experimental group mean judgements and control group mean judgements under the two conditions. The shifts on the whole appear to be slightly larger in condition 2,

suggesting that the labelling effect may be related either to the confusability or to the relative unfamiliarity of the stimuli used. This point will be discussed in greater detail later.

The direction of the shifts in relative similarity judgements shown in the graphs, suggests three hypotheses that may be fruitful to examine regarding the effect of labels on taste similarity judgements.

Hypothesis 1 takes account of the fact that there is a preponderance of control group points which exceed in magnitude (on the similarity / dissimilarity scale) the appropriate experimental group points for each given stimulus pair. The hypothesis simply states that nonsense labels reduce the judged similarity between taste stimuli as compared with the judgements produced when subjects judge the same taste stimuli but use their own preferred labels. In fact, the graphs and associated tables show that 15 of the 26 pairs of points (57.7%) in condition 1, and 17 of the 25 pairs (68.0%) in condition 2 differ in the direction predicted by this hypothesis. (There is one pair of tied points in condition 2 which will be ignored in this discussion.)

A more powerful hypothesis can be stated, however. A closer examination of the distribution of points across the similarity scale suggests that there is a tendency for the control group mean judgements to be more extreme towards both poles of the similarity scale. To evaluate this hypothesis it is useful to take some kind of mid-scale baseline. We then predict that when the control group judgements have a value which lies above the baseline, the

experimental group judgements will be smaller in magnitude on the 10 point similarity scale. However, when the control group judgements have a mean value that is less than the magnitude of the baseline, then the experimental judgements will be greater than the control group on the similarity scale. This approach is not unlike the unfolding theory of Coombs (1964) in that we are interested in distances between pairs of points; one point represents the individual's "ideal point" (the "baseline" in our terminology) while the other point represents the stimulus in question. We would expect the actual value of the ideal point or baseline to vary from subject to subject, however for the present experiment it may be adequate to assume that the individual baselines will vary only slightly so that analysis of pooled data may reflect the underlying trends.

There are two suggestions for determining the actual magnitude of the baseline. We can take the grand mean similarity, that is, the mean of the entire set of judgements made by the control group, as the measure of the mid-point or central tendency of their judgements using the scale of similarity. Alternatively, and more simply, we could take the natural midpoint of the scale, i.e. a scale value of 5.0 as our baseline. For condition 1, the grand mean similarity for the 5 control subjects was 6.089; for condition 2 it was 5.961. Using these grand mean similarities as our baselines, we find that 18 out of 26 or 69.2% of the observations in condition 1 and 19 out of 25 or 76.0% of the points in condition 2 fit the second hypothesis. (By using the grand mean similarity as the value for our baseline we allow for the possibility

that the overall scale usage may be biased towards one end or the other of the 10 point similarity scale.)

Hypothesis 3 uses the natural midpoint of 5.0 as the baseline. In this case we get a better fit with 19 out of 26 or 73.1% of the pairs of points in condition 1 and 21 out of 25 or 84.0% of the pairs in condition 2 lying in the direction predicted by the hypotheses. Note that hypotheses 2 and 3 do not predict that the judgements produced by the experimental groups will always be intermediate between the mean judgements for control group subjects and the baseline. This more rigid hypothesis is not well substantiated by the data. Rather, hypotheses 2 and 3 can be interpreted as follows:

Let  $j, k$  represent a pair of stimuli to be judged against each other.

$C$  is the mean judged similarity produced by the control group.

$E$  " " " " " " " " experimental group.

Then the two hypotheses state simply that

If  $C_{jk} > \text{baseline}$ , then  $E_{jk} < C_{jk}$

If  $C_{jk} < \text{baseline}$ , then  $E_{jk} > C_{jk}$ .

This model will be further developed in later chapters.

Assuming that the baseline hypotheses do in fact represent the true state of affairs, what psychological interpretation can we place on this finding? It appears that one of the effects of using nonsense labels (as distinct from subject's own labels) is that the nonsense labels somehow reduce the perceived or judged distinctiveness of the tastes, as shown by a reduction in the range

of scale used by the experimental group subjects. (In fact, an additional control group in which no labels were paired overtly with the tastes would have been useful.) Perhaps the experimental treatment effect can be attributed to some kind of strategy whereby subjects given labelled stimuli judge similarity not in terms of the physical stimuli alone, but in terms of the total complex, stimulus (taste) plus label. Since the experimental group, in a sense, judged a set of tastes with uniformly meaningless labels (subject to errors in the word association measures), while the control group used relatively meaningful names, we would predict that the experimental judgements would show a smaller range of usage on the similarity continuum, in much the same way as we would expect the 2 dimensional stimuli red square, red circle, and red triangle to produce a less extreme set of judgements than say, a red square, a blue circle and a green triangle.

The suggestion that subjects handle the labelled tastes as 2-dimensional also gains some support from some results discussed earlier; namely, the observation that the judgements made by subjects in condition 2 showed a greater shift due to labelling than the judgements made by subjects in condition 1. The suggestion is that when discrimination between the tastes is more difficult, labelling becomes a more significant factor in the judgement task. (If we consider the labels as a second dimension, then in condition 2, as the gustatory dimension becomes more confused, greater emphasis is placed on the verbal dimension by subjects, hence producing a greater separation of points for the judgements

made by experimental, as opposed to control group subjects.)

So far we have not considered the question of whether or not the trends noted in the data can be shown to be statistically significant. The major part of the analysis for this experiment consisted of a series of discriminant function analyses, which will now be discussed in some detail.

The discriminant function was introduced by R.A. Fisher in 1936 as a statistical technique to facilitate the classification of objects or persons. The basic requirements are that each subject generates a set of measurements and that we have related sets of measures on the same variables in each of the alternative groups into which we wish to classify the individual subjects. The problem is to assign the subject to one or the other of these groups by appropriately weighting these measures to produce the classification associated with the smallest probability of error. In terms of the present experiment, given that we have 4 treatment groups, each with its own characteristic profile of mean judgements that is based on the pooled data from the 5 subjects in that group, what are the statistical probabilities that any one subject will be classified as belonging to one or the other of the 4 groups? A priori, we know to which treatment group the subject actually belongs, but the problem is to show statistically that his response profile does in fact indicate his membership within that group, rather than within any of the 3 remaining groups. The statistical procedure for calculating the probabilities that a given subject is associated with each of the treatment groups is outlined in general



terms in Hope (1968) and a worked example is discussed in Porebski (1966).

The analyses reported here were calculated on the I.B.M. 360/44 Computer using a program based on that developed by Dixon (1964) and Anderson (1958) and modified by Gregson (1968) to give the Bayesian a posteriori probabilities associated with the membership of each subject in each of the treatment groups. It is necessary that the total number of subjects minus two should not be less than the number of scores or variables (in this case we have  $8 \times 4 = 32$  variables or similarity measures), otherwise a trivially perfect separation of groups is produced. For this reason the data were broken down into smaller blocks of 4 variables and 16 separate analyses were carried out; 8 analyses on the raw or nonstandardised data and 8 analyses on the same data in standardised form.

In analyses 1 to 4, the 4 variables in each case were made up by combining, for each taste in turn, the simple standards judged against the mixtures along with the mixture standards judged against the simple tastes. Thus analysis 1 was based on the following matrix format:

		Comparison Stimuli			
		Sucrose	NaCl	Citric Acid	Quinine Sulphate
Sucrose Std.	Condition 1 Exptal.				
	Condition 1 Contr.				
-----					
Suc. Mixt. Std.	Condition 2 Exptal.				
	Condition 2 Contr.				
		Suc.	NaCl	Citr. A.	Quin. S
		Mixt.	Mixt.	Mixt.	Mixt.
		Comparison Stimuli			

Analyses 2, 3 and 4 followed the same format except that the standards were sodium chloride, citric acid, and quinine sulphate simple and mixture standards. Analyses 1 to 4 will be referred to as the simple / simple and mixture / mixture analyses.

Analysis 5 consisted of a different combination of variables:

Comparison Stimuli

			Sucrose	NaCl	Citric Acid	Quinine Sulphate
			Mixt.	Mixt.	Mixt.	Mixt.
Sucrose Std.	Condition 1	Exptal.				
	Condition 1	Contr.				
-----			-----	-----	-----	-----
Suc. Mixt. Std.	Condition 2	Exptal.				
	Condition 2	Contr.				
			Sucrose	NaCl	Citric Acid	Quinine Sulphate

Comparison Stimuli

Analyses 6, 7 and 8 followed the same format except that the standards were sodium chloride, citric acid and quinine sulphate simple standards and mixture standards.

The 8 analyses were then repeated on the standardised data, using the same combinations of variables as in the 8 raw data analyses. The results of the 16 analyses appear in Table 2:5. In all 16 analyses, the Mahalanobis D squared is distributed as chi-squared with 12 degrees of freedom; it provides an indication of the chance level for which we would expect to find the calculated probabilities associated with membership of a given subject in each of the 4 experimental treatment groups.

TABLE 2:5

Simple / simple and mixture / mixture combined analyses:

<u>Raw data</u>			<u>Standardised data</u>		
<u>Analysis.</u>	<u>Mahalanobis</u> <u>D squared.</u>	<u>Signif.</u> <u>level.</u>	<u>Analysis.</u>	<u>Mahalanobis</u> <u>D squared.</u>	<u>Signif.</u> <u>level.</u>
1.(Sucrose)	245.29	0.001	5.(Sucrose)	129.64	0.001
2.(NaCl)	63.13	0.001	6.(NaCl)	75.62	0.001
3.(Citr.A.)	64.09	0.001	7.(Citr.A.)	58.02	0.001
4.(Quin.)	53.21	0.001	8.(Quin.)	68.85	0.001

Simple / mixture and mixture / simple combined analyses:

<u>Raw data</u>			<u>Standardised data</u>		
<u>Analysis.</u>	<u>Mahalanobis</u> <u>D squared.</u>	<u>Signif.</u> <u>level.</u>	<u>Analysis.</u>	<u>Mahalanobis</u> <u>D squared.</u>	<u>Signif.</u> <u>level.</u>
9.(Sucrose)	32.19	0.01	13.(Sucrose)	18.82	N.S.
10.(NaCl)	9.15	N.S.	14.(NaCl)	14.84	N.S.
11.(Citr.A.)	18.06	N.S.	15.(Citr.A.)	28.92	0.01
12.(Quin.)	7.32	N.S.	16.(Quin.)	12.19	N.S.

In general, the transformation of the raw data into standardised scores appears to have had little effect on the results of the analyses. The following discussion will therefore be limited to the results of the analyses of the raw data.

It is readily apparent from Table 2:5 that the simple / simple and mixture / mixture analyses produced highly significant results, while the simple / mixture analyses tended to be nonsignificant statistically. There is some evidence that the experimental task

became too difficult when subjects were required to judge relatively familiar standards against unfamiliar comparison stimuli. The discrimination process appears to have broken down with a corresponding increase in the error variance of the judgements made. This breakdown in discrimination can also be seen by referring to the frequencies with which comparison stimuli were apparently incorrectly identified. Since, for each subject, 4 of the comparison stimuli are physically identical to the 4 standards, while the other 4 are closely related to the standards it seems reasonable to assume that for a perfectly discriminating subject the physically most closely related stimuli should be judged as most similar. For example, a subject in condition 1, given sucrose as a comparison stimulus would be expected to judge it as most like standard 1 (sucrose). If some other standard (2, 3, or 4) was judged as most like sucrose, we can infer that the subject has misidentified the comparison stimulus. By this definition of misidentification, the following judgements between physically identical standards and comparison stimuli were observed to be based on incorrect identification of the comparison stimulus:

<u>Stimulus pair</u>	<u>Frequency of misidentification</u>
Sucrose / Sucrose	0
NaCl / NaCl	0
Citric Acid / Citric Acid	4
Quinine Sulphate / Quinine Sulphate	7
Sucrose Mixt. / Sucrose Mixt.	9
NaCl Mixt. / NaCl Mixt.	19

Citric Acid Mixt. / Citric Acid Mixt.	16
Quinine Sulphate Mixt. / Quinine Sulphate Mixt.	16

By a similar argument we would expect that sucrose would be judged as more similar to sucrose mixt. standard (and vice versa) than to any of the other 3 mixture standards. In this case the frequencies of misidentification were as listed below (the first member of each pair is the standard; the second is the comparison stimulus):

<u>Stimulus pair</u>	<u>Frequency of misidentification</u>
Sucrose / sucrose Mixt.	13
NaCl / NaCl Mixt.	8
Citric Acid / Citric Acid Mixt.	19
Quinine S. / Quinine S. Mixt.	14
Sucrose Mixt. / Sucrose	2
NaCl Mixt. / NaCl	8
Citric Acid Mixt. / Citric Acid	17
Quinine S. Mixt. / Quinine S.	10

Overall, the total number of misidentifications for judgements between a familiar standard, and the relatively unfamiliar comparison stimuli which were not physically identical to the standards, totalled 91, while the misidentifications between familiar standards and identical comparison stimuli totalled 71, suggesting greater confusion in the simple / mixture judgements than in the simple / simple and mixture / mixture judgements. This may account for the failure of the former pairs of judgements to produce statistically significant results.

We shall disregard the nonsignificant analyses and concentrate on the results from analyses 1 to 4.

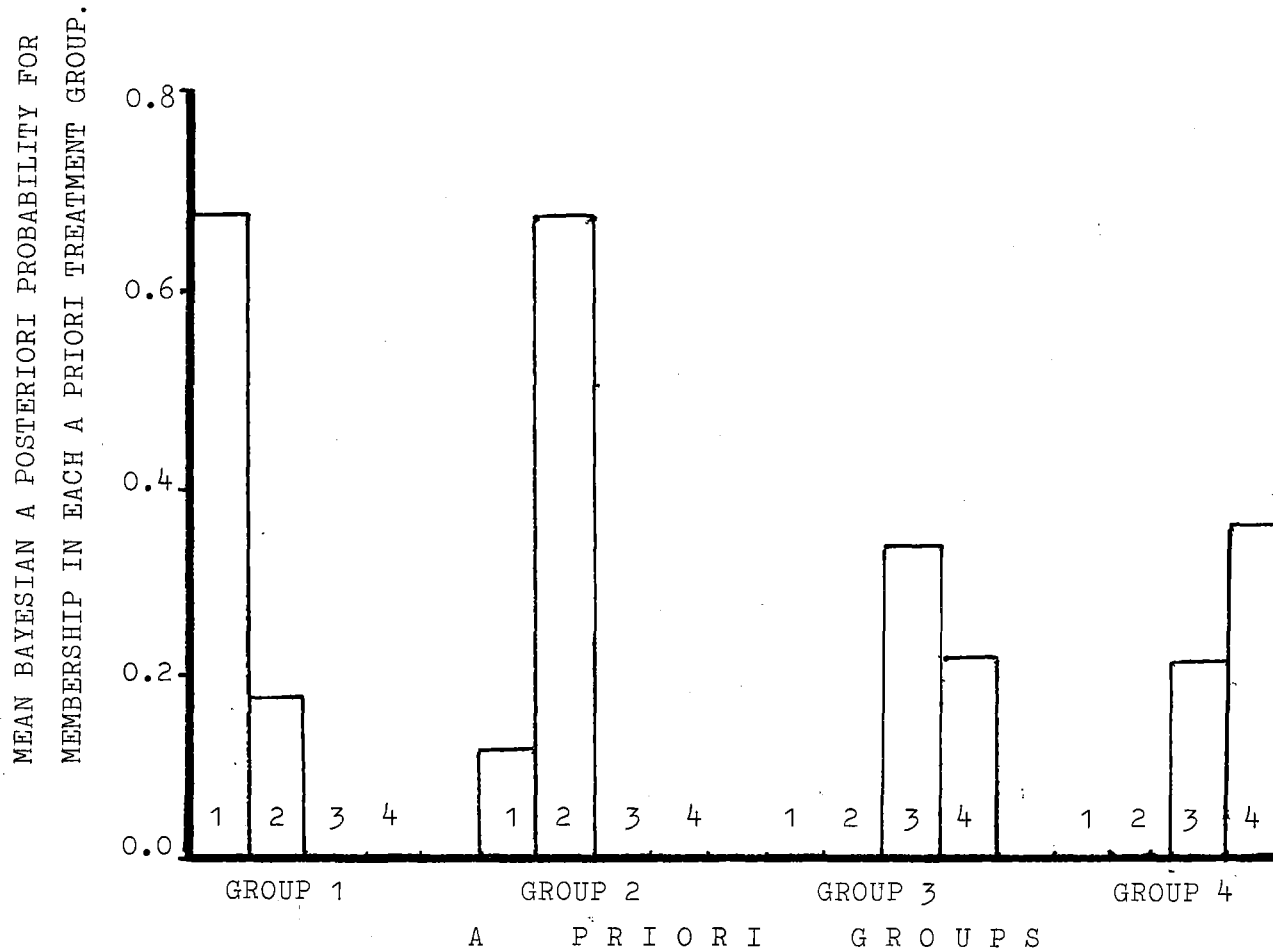
Although the analyses were carried out at the individual subject level the data and results from the five subjects in each of the four treatments have been pooled and averaged to simplify this discussion. The discriminant function analysis produces two sets of measures which show the probability with which a given subject can be assigned to each of the four experimental treatment groups on the basis of his particular set of taste similarity judgements. These two sets of measures are (a) the relative likelihoods and (b) the Bayesian a posteriori probabilities of group memberships.

Since there are four treatment groups, we would expect, a priori, if the experimental treatment had no effect on the judgements generated by each subject, that each subject would be assigned to each group with a probability of 0.25. If only one of the two factors under investigation had any influence on the similarity judgements given by subjects, then subjects would be approximately assigned to two of the treatment groups with a probability of 0.5 and to the other two groups with a probability of 0.0. In this way, the probabilities give us an approximate indication of the relative importance of the two factors under investigation here.

Table 2:6 has been condensed from analyses 1 to 4 by taking the 5 subjects belonging to each of the 4 a priori experimental groups and averaging their a posteriori group membership scores to give the

FIGURE 2:3

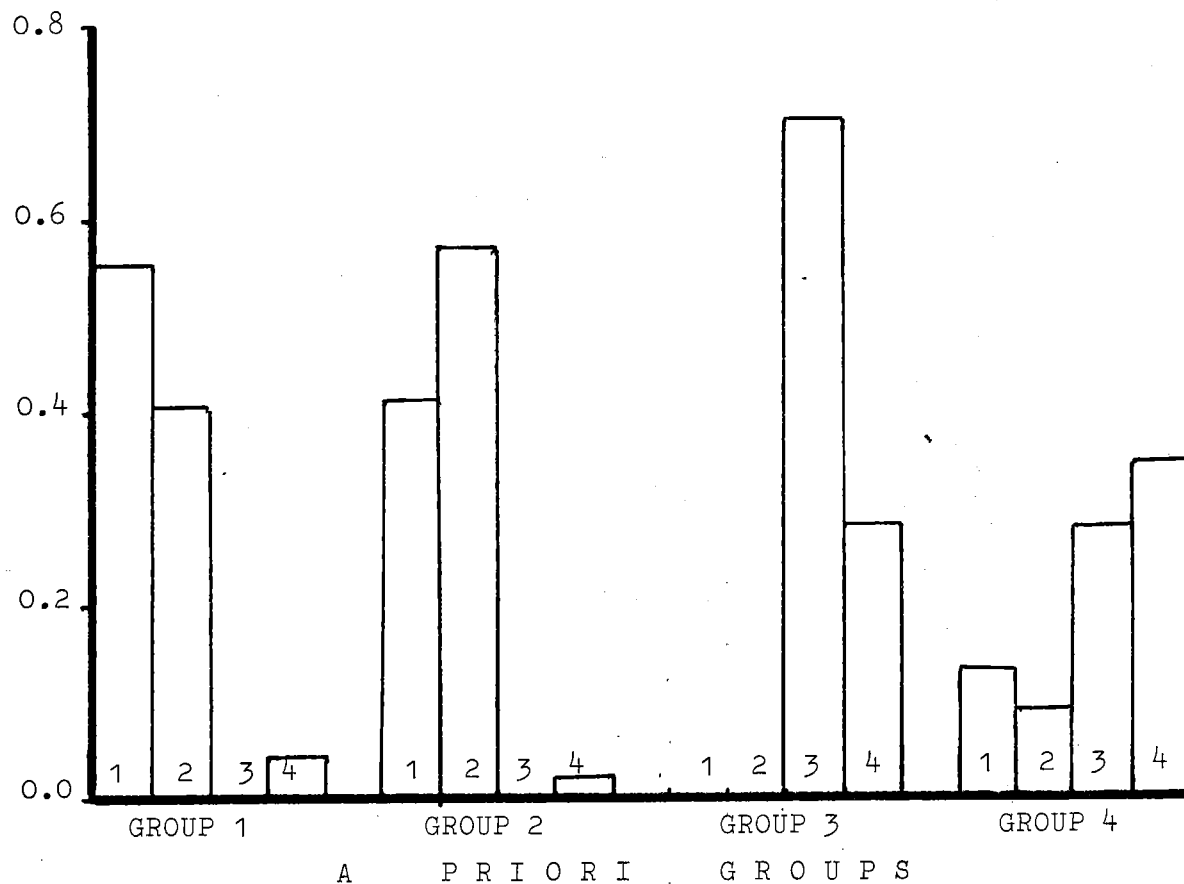
SUCROSE SIMPLE WITH SIMPLE  
AND MIXTURE WITH MIXTURE  
COMBINED ANALYSIS.



MEAN BAYESIAN A POSTERIORI PROBABILITY FOR  
MEMBERSHIP IN EACH A PRIORI TREATMENT GROUP.

FIGURE 2:3

SODIUM CHLORIDE      SIMPLE WITH SIMPLE  
AND MIXTURE WITH MIXTURE  
COMBINED ANALYSIS.





MEAN BAYESIAN A POSTERIORI PROBABILITY FOR  
MEMBERSHIP IN EACH A PRIORI TREATMENT GROUP.

FIGURE 2:3

CITRIC ACID      SIMPLE WITH SIMPLE  
AND MIXTURE WITH MIXTURE  
COMBINED ANALYSIS.

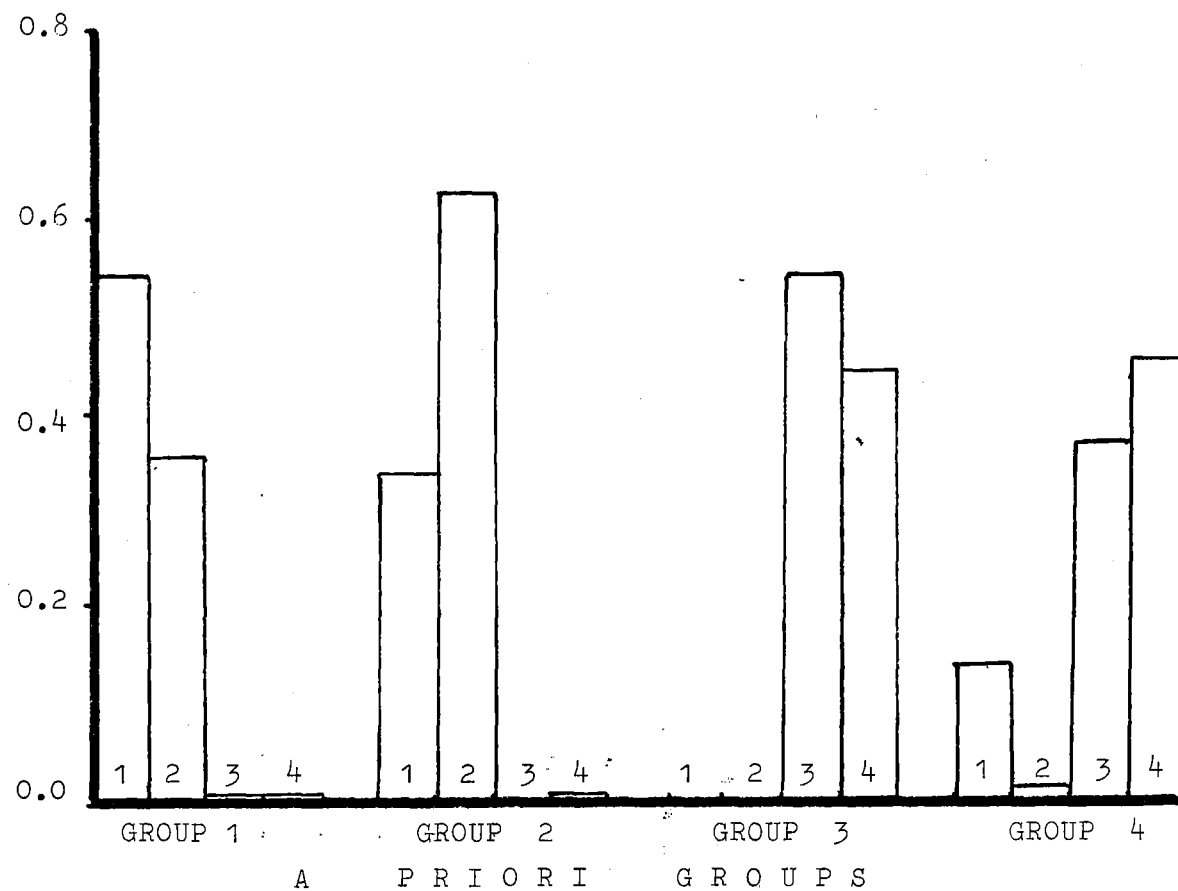


FIGURE 2:3

QUININE SULPHATE      SIMPLE WITH SIMPLE  
AND MIXTURE WITH MIXTURE  
COMBINED ANALYSIS.

MEAN BAYESIAN A POSTERIORI PROBABILITY FOR  
MEMBERSHIP IN EACH A PRIORI TREATMENT GROUP.

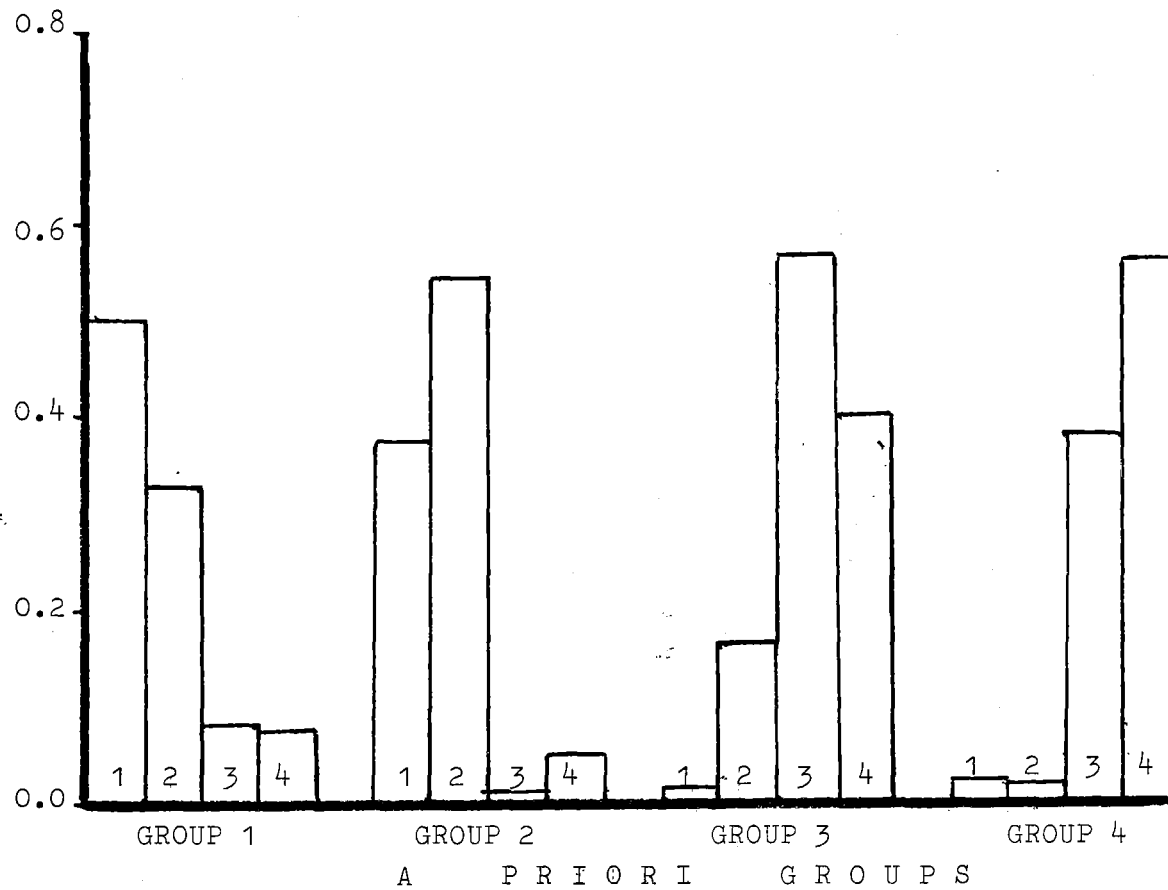


TABLE 2:6

MEAN BAYESIAN A POSTERIORI PROBABILITIES OF MEMBERSHIP IN EACH  
TREATMENT GROUP POOLED OVER THE 5 SUBJECTS FROM EACH A PRIORI

TREATMENT GROUP.

ANALYSIS	A PRIORI TREATMENT UNDER WHICH SUBJECTS RESPONDED							
	T1	(Cond. 1 Exptal.)	T2	(Cond. 1 Control)	T3	(Cond. 2 Exptal.)	T4	(Cond. 2 Control)
1 (Sucrose)	T1	0.6686		0.1131		0.0000		0.0000
	T2	0.1693		0.6668		0.0000		0.0000
	T3	0.0000		0.0000		0.3305		0.2066
	T4	0.0000		0.0000		0.2099		0.3272
2 (NaCl)	T1	0.5488		0.4075		0.0089		0.1359
	T2	0.4076		0.5725		0.0046		0.0917
	T3	0.0007		0.0001		0.7062		0.2818
	T4	0.0429		0.0202		0.2802		0.4906
3 (Citr.A.)	T1	0.5032		0.3610		0.0065		0.1424
	T2	0.3552		0.6343		0.0002		0.0191
	T3	0.7215		0.0016		0.5446		0.3785
	T4	0.6938		0.0031		0.4487		0.4600
4 (Quin.S.)	T1	0.5031		0.3772		0.0134		0.0273
	T2	0.3316		0.5446		0.1708		0.0248
	T3	0.0848		0.0163		0.5768		0.3857
	T4	0.0804		0.0556		0.4060		0.5621

mean probabilities for membership in each group. (These mean probabilities are shown in the table.) The mean Bayesian a posteriori probabilities shown in Table 2:6 are plotted as a series of histograms in Figure 2:3. In all 16 of the histograms there is a 1 to 1 matching between the a priori and the a posteriori group memberships. That is, in all cases the largest a posteriori probability is associated with the actual experimental condition under which the 5 subjects in that group were treated.

See Table 2:6.

A second point to emerge is that the partitioning between treatment groups 1 and 2 as opposed to 3 and 4 is much more clearly defined than the partitioning between groups 1 and 3 vs. 2 and 4. This suggests that the first treatment factor (simple vs. mixture standards) is a far stronger treatment effect than the second factor (the labelling treatment). In fact, the statistically significant results obtained may be largely attributable to the first factor in each case - the labelling effect is quite definitely second order in this experiment.

The most clear-cut partitioning on the labelling factor appears in analysis 1 (sucrose simple / simple and mixture / mixture). Keeping this observation in mind, we will now return to the question of "overwriting" discussed earlier in this chapter.

Let us state, as an initial assumption, that the subject's own verbal schema under which he classifies a set of taste qualities can be overwritten by the experimenter to produce systematic shifts in judgements. The data analysed to date do suggest that such an

effect has been produced in the present experiment. We will assume then, for the present argument, that overwriting has occurred; that is, we assume that subjects establish verbalised conceptual schemata of taste qualities during their day-to-day interactions with taste stimuli in their environment. We can further postulate that the more commonly experienced, or more familiar tastes are more likely to have a label in the subject's schema than the less familiar tastes, i.e.

$$\begin{aligned} &\text{Pr. (that subject has a label for a given taste)} \\ &= f(\text{prior exposure to that taste}). \end{aligned}$$

The above hypothesis can be restated thus:

$$\begin{aligned} &\text{Pr. (that subject's own label can be overwritten)} \\ &= f(1 / \text{prior exposure to the labelled taste}). \end{aligned}$$

In other words, overwriting is more likely to be achieved with an unfamiliar taste like quinine than with a familiar taste like sucrose. However, the point mentioned earlier, that the influence of labelling is most marked in the case of sucrose, suggests that the hypothesis relating ease of overwriting to familiarity of the stimulus is not an adequate conceptualisation of the labelling phenomenon. On the basis of prior exposure we would expect (at least in subjects from typical Western cultures) that sucrose would be the most familiar taste quality, (or at least the most extremely overlearned labelled taste sensation, matched only by "salty"), and indeed the frequencies of misidentifications of standards listed earlier tend to support this premise. On the other hand, it also follows from the familiarity hypothesis that the simple tastes are more likely to have

already-established labels or names than the mixtures, which are very unlikely to have been experienced previously. On this basis the familiarity hypothesis would predict that the condition 2 data would show a greater response shift due to labelling than the data from condition 1, and although this tendency does appear to a slight extent, the increase in noise or error variance in condition 2 makes any such interpretation difficult.

An alternative approach, which may lead to a clearer understanding of the phenomena underlying the present experiment is to examine the data in terms of symmetry. If  $s$  represents the judged similarity between stimuli  $x$  and  $y$ , then the symmetry principle simply states that

$$x \text{ } s \text{ } y \quad \equiv \quad y \text{ } s \text{ } x$$

In the present experiment it is possible to compare for symmetry, judgements made between labelled standards and unlabelled comparison stimuli only; in no cases were the comparison stimuli labelled. It may be of value to explore the following hypothesis in terms of symmetry of the data. The hypothesis examines the proposition that symmetry may be related to the labelling factor and predicts that greater asymmetry will be shown under one labelling condition than under the other.

First, however, some underlying predictions and postulates will be developed. Let us consider the stimuli in terms of the 'engrams' or memory traces that we might expect them to produce within subjects. We can postulate that memory traces differ in the degree to which they are differentiated within subjects and that the

differentiation relates to a number of factors including the following:

- (1) The degree to which the individual subject has been previously exposed to the stimulus concerned is directly related to the extent to which the internal representation or memory trace of the stimulus is differentiated.
- (2) The degree of differentiation of the memory trace decreases as the time interval since the stimulus was last encountered increases. (This postulated decay in the memory trace is considered to occur over a matter of minutes; it has been empirically tested by Brown (1958) with positive results.)
- (3) Shepard (1957) has developed the postulate that stimulus confusions and response confusions occur independently of each other, and that the probability of both occurring can be described as an exponential decay function of the psychological distance between the stimuli or responses concerned.
- (4) If a label which adequately conceptualises or describes relevant aspects of the stimulus is associated by the subject with the stimulus, then this label will reduce the rate of decay of the memory trace over time. In the present experiment, the judgements involve the comparison between a standard, which is carried in the memory, and a comparison stimulus (with its associated receptor input), which is actually held in the mouth while the judgement is made.

We can now make a number of specific predictions:

- (a) Real-word labels will enhance to a greater extent the

distinctiveness of relatively unfamiliar (undifferentiated) taste qualities, than the distinctiveness of familiar tastes which are already well differentiated due to more frequent day-to-day experience of the taste.

(b) Nonsense labels will reduce the distinctiveness of tastes by overwriting their real-word labels with a meaningless "description" or "conceptualisation".

We will now examine these two predictions in terms of the symmetry principle.

Let A and B represent a pair of taste stimuli; s represents the judged similarity between them.

If we assume that symmetry holds in the absence of experimental manipulation then statement (a) above predicts that the pairs

(i) A (labelled) s B (unlabelled)  
and (ii) B (labelled) s A (unlabelled)  
are unequal or asymmetrical.

Let us further assume that A is a familiar taste and B is an unfamiliar taste stimulus, and also that when a subject cannot discriminate between two taste stimuli he will judge them to be highly similar. Now if, as we predict, labelling exerts a greater influence on unfamiliar stimuli then pair (i) above will be less precisely discriminated from each other (i.e. judged to have greater similarity) than pair (ii), since the unfamiliar stimulus is not labelled in (i). Thus we predict that a familiar taste judged against a less familiar taste will be judged lower (nearer the "identical" pole) on the scale than its symmetrically opposite pair,



when the first taste of each pair is the one which has been overtly paired with a real word label.

This prediction was tested as follows. The four simple standards were arranged in a hierarchy of decreasing familiarity which was based on the number of misidentifications of each taste mentioned earlier, i.e. sucrose  $\Rightarrow$  sodium chloride  $\Rightarrow$  citric acid  $\Rightarrow$  quinine sulphate. The symmetrically opposite pairs of stimuli and the responses given to them by each of the 5 subjects in the 4 experimental conditions were tabulated; the results for the 10 subjects from condition 1 only are shown in Table 2:7.

Each column of the table is headed by a pair of taste stimuli. The first entry in each cell is the mean judgement made by a subject when the first of the pair of taste stimuli heading the column was the standard, while the second entry in each cell (in parentheses) is the mean judgement made when the second stimulus served as the labelled standard. Since the order in which the pairs are listed follows our hierarchy of familiarity, our predictions are that the first entry in each cell of the table will be of smaller magnitude than the second entry in the cell for the control subjects, since they used real labels. For the experimental subjects, however, we expect the opposite to be true; the nonsense labels hinder discrimination rather than enhance differentiation of the memory trace, and so we predict that the asymmetry will lie in the opposite direction (i.e. the first entry in each cell will be of larger magnitude than the second.) In fact, if we neglect tied values, a total of 23 out of 28, or 82.1% of the points differ

TABLE 2:7

	<u>Suc.Std./NaCl</u>	<u>Suc.Std./Acid</u>	<u>Suc.Std./Quin.</u>	<u>NaCl Std./Acid</u>	<u>NaCl Std./Quin.</u>	<u>Acid Std./Quin</u>
Subjects	1. 9.75 (9.5 )	9.25 (8.25)	10.0 (10.0)	5.75 (6.0 )	7.0 (8.75)	7.25 (9.0 )
	2. 7.5 (5.5)	8.75 (8.25)	10.0 (9.75)	5.75 (5.5 )	8.5 (7.0 )	2.75 (2.0 )
	3. 7.5 (7.75)	10.0 (9.75)	10.0 (9.25)	7.5 (6.5 )	7.75 (6.0 )	6.25 (2.5 )
	4. 7.5 (7.5 )	8.0 (7.5 )	10.0 (9.25)	7.5 (7.0 )	8.0 (7.5 )	7.0 (7.5 )
	5. 8.25 (8.0 )	8.75 (6.75)	8.5 (6.75)	6.75 (6.0 )	7.25 (6.25)	5.0 (4.0 )
	6. 10.0 (8.0 )	8.5 (8.5 )	9.0 (9.0 )	5.75 (7.5 )	8.0 (9.0 )	6.5 (8.5 )
Subjects	7. 9.25 (8.75)	9.25 (8.25)	8.25 (9.5)	6.75 (6.0 )	6.25 (8.0 )	7.75 (4.0 )
	8. 8.75 (9.25)	9.5 (8.25)	9.75 (9.75)	5.0 (3.75)	5.5 (6.75)	6.5 (4.75)
	9. 7.5 (7.25)	8.0 (8.25)	8.25 (8.0)	8.5 (8.25)	7.75 (7.75)	8.0 (7.25)
	10. 8.5 (8.25)	9.75 (10.0)	8.75 (5.25)	4.0 (4.25)	6.5 (7.5 )	2.75 (7.5 )

in the predicted direction for the experimental subjects while only 12 out of 26, or 46.1% of the points lie in the predicted direction for the control group subjects. For the ten subjects in condition 2, the proportions were 13 out of 27 (48.1%) for the experimental group and 16 out of 28 (57.1%) for the control group. Of these values only the first, for the condition 1 experimental group, is statistically significant. The Binomial test produced a z-value of -3.212; the two tailed probability of gaining a value equal to or greater than this is 0.0014.

The hypotheses have only been subjected to weak tests here however. The failure of the other 3 treatment groups to reach statistical significance means only that the predicted labelling effect did not necessarily exceed the suggested influence of taste familiarity. Our aim at this point is merely to exploit the data as a means of suggesting new directions for future research.

This chapter will now be concluded with a summary of the main findings and suggestions for future research.

Two distinct theoretical approaches have been suggested. The first, a trace decay model, assumes that our observation that different verbal labelling schemata can produce reliable shifts in subjective similarity judgements, can be attributed to a slowing down of trace decay caused by the pairing of subject-provided real word labels for the stimuli. It could be argued that when we pair nonsense labels with tastes, the uniformly meaningless cue provided by the labels increases the generalization between pairs of stimuli and this increase in generalization produces the decrease in range

of scale values used in assessing taste similarities. The finding that labelling may induce asymmetrical responses to physically symmetrical pairs of stimuli can be attributed to the facts that

(a) we would expect trace decay to be more rapid for unfamiliar taste stimuli than for familiar taste stimuli, and that

(b) nonsense labels would be expected to have a greater effect on reducing the differentiation of unfamiliar than of familiar taste sensations.

The second approach assumes that subjects treat labelled tastes as multidimensional stimuli and weight both physical taste and label as separate elements when coming to an overall assessment of similarity. Since the nonsense labels can be considered to add a set of uniformly meaningless elements to each stimulus we find a decrease in the range of scale usage when nonsense labels are paired with the taste stimuli. The finding that labelling induces asymmetry in physically similar pairs of stimuli can again be attributed to the fact that nonsense labels add perceptually similar cues to potentially dissimilar stimuli, but the effect of these labels on increasing the generalization between tastes is greater for unfamiliar than for familiar stimuli. We cannot distinguish between the validity of these two theoretical approaches with the present data. However, no matter which explanation we accept, it appears that by pairing unfamiliar or unusual names with stimuli we can "overwrite" the existing labels which are presumed to be used covertly by subjects to refer to taste qualities. Secondly, this "overwriting" can be empirically demonstrated to induce shifts

in the judged similarity between labelled and unlabelled stimuli. The results of this pilot study suggest that words play a small, but active part in the judgement process of human subjects; that stimulus labels can, in a sense, carry additional information which supplements the physical input arising from stimuli, and that this additional information influences the outcome of subjective similarity assessments. These points will be examined in greater detail in the chapters which follow.

### CHAPTER III

#### THE EFFECT OF NONSENSE LABELS ON THE

#### JUDGED SIMILARITY OF RECTANGLES.

In the pilot study, there was some suggestion that stimulus labels can play a part in determining the outcome of the judged similarity between pairs of stimuli. One explanation advanced was that the nonsense labels which were used were similar in "meaning" (or perhaps uniform in their lack of meaning) and consequently they reduced the degree to which differences in the actual physical stimuli were weighted by subjects during the judgement task. The labels thus reduced the range of responses generated by those subjects for whom nonsense labels had been paired with the taste stimuli by the experimenter. An alternative explanation, based on trace decay theory, was also put forward.

However, a number of important points were not adequately accounted for in the experimental design:

1. There was no control over the temporal delay between presentation of the standard and presentation of the comparison stimulus which was to be judged against it. For example, presentation of the first comparison stimulus in Part 3 of the experiment, was separated from Part 2 in which the standards had been presented and labelled, by a maximum interval of 3 minutes, while presentation of the last comparison stimulus followed presentation of the standards by almost half an hour.

2. There was no control over the possible degree to which the memory traces associated with each of the 4 standard taste sensations may have decayed progressively over trials in Part 3 since the standards were only presented and labelled as such in Part 2.

3. Of the total set of 8 taste stimuli, the 4 standards only were overtly labelled by the experimenter. There was no control over the covert labels which subjects may have assigned to the comparison stimuli during Part 3 of the experiment. Furthermore, we cannot assume that a subject who had learned to identify 4 tastes by using 4 nonsense syllable names during Part 2 of the experiment actually used the same names when these 4 tastes appeared as comparison stimuli in Part 3 of the experiment.

4. Whenever possible, it is desirable to use each subject as his own control, rather than to compare groups of subjects as was necessary in the pilot study.

The above considerations have made it difficult to formulate specific predictions about the way in which the judged similarity between pairs of stimuli would be expected to shift under the influence of the manipulation of labels.

It is of considerable theoretical interest to know whether the labelling effect observed for taste is found in other sense modalities, particularly in the senses which are more dominant in the human, such as vision and hearing. There are advantages in using the sense of taste as a model sensory system, as Pfaffmann (1962) has pointed out. However, in the present experimental approach, additional disadvantages arise from the fact that the

stimuli must be presented sequentially, preferably with a fresh-water rinse between stimuli, and with a time interval to allow for recovery from receptor fatigue and adaptation.

It was therefore decided to conduct the following experiment using visually presented stimuli, with a minimum period of delay between presentation of the standard and comparison stimuli, and with all stimuli labelled in the experimental condition.

One point which does not appear to have received much attention in the literature is the relationship between the cognitive processes involved in similarity judgement and the processes involved in typical concept formation tasks. Nurminen (1965) has pointed out that if the perceived relationship (e.g. similarity) between stimuli can be changed by pairing names with the stimuli then this will have relevance to the study of concept formation. Part of the literature reviewed in Chapter I has concentrated on the role of stimulus labelling in concept formation. For example, Fenn and Goss (1957) found that labels paired with blocks influenced the sorting of blocks into sets, such that, in a sense, the perceived similarity in terms of certain relevant dimensions was influenced by labelling. It may be of some relevance, therefore, to consider whether similarity should be regarded as a concept or as a process (such as generalization). Wallach (1958) has discussed four ways in which psychological similarity has been defined. The first approach is simply to define similarity in terms of common environmental properties. This approach makes the invalid assumption that if common properties are present in the



environment they will be perceived. A second approach aims at a more operational definition by defining similarity in terms of common responses - if a person responds in the same way to two objectively different situations then the two situations are psychologically similar for him. For example, in the Shipley (1935) experiment the common response of winking made to both light and shock indicated their psychological similarity for the subject. Osgood (1953) and others developed this view that the common response itself constitutes psychological similarity - this approach has been discussed in Chapter I. Yet another approach discussed by Wallach is to suggest that neural traces are laid down along various dimensions when a stimulus impinges on the organism and psychological similarity depends on how far a new stimulus is located from the old stimulus on such a dimension. This is the theory of primary stimulation gradients; it is particularly suited to describing those cases of similarity that are based on generalization and discrimination learning. The fourth approach and the one which Wallach finds the most fruitful is to conceive of similarity as dependent on the application of a rule which leads one to assign items to a common category. Prior to making the assignment, the subject must, of course, learn the rule which guides the response. To quote Wallach, ".... recognition of the similarity of two events depends on their being classed as equivalent ...." (p.112). He adds that the classification rules seem to be influenced by two main determinants: (a) the nature of the instances presented (i.e. the environmental properties to which we have been exposed) and

(b) our predilections for some bases of classification and prejudices against others.

This fourth view of similarity as proposed by Wallach bears a striking resemblance to the methods by which subjects are considered to identify concepts. Kendler (1961) for example, has defined concept learning as any situation that involves the "acquisition or utilization or both of a common identifying response to dissimilar stimuli". Hunt (1962), on the other hand, considers concept learning to be the learning of names, where a name is a symbol used to refer to a set; the concept is thus the meaning of the name (Church, 1958). Wallach (1958) further argues that by varying the basis on which we classify stimuli it should be possible to influence psychological similarity in predictable ways. That is, by facilitating or hindering the ease with which subjects can arrive at rules for classifying stimuli in terms of properties which they potentially have in common, we should be able to influence judged similarity. One method of altering the potential classification of a set of stimuli is by differentially pairing verbal labels with individual stimuli. According to Wallach's fourth definition of similarity, then, if we can effectively alter the subject's classification of the stimuli, we should also be able to manipulate the judged similarity of the stimuli as they appear under the altered classification scheme. This point will be discussed further in the concluding pages of this chapter.

Early studies by Hull (1939) have shown that generalization gradients which are comparable to those obtained in classical

conditioning studies can be obtained by using appropriate instructions and scaling procedures which are not unlike the scale of similarity / dissimilarity used here. A necessary factor is that the stimuli on the continuum to which the subjects are to respond should be physically separated by equal intervals.

It was proposed that the following experiment should investigate whether or not judged similarity can be manipulated by using names to reclassify a set of stimuli into new classes or conceptualisations which differ from the ways in which the stimuli would be expected to be classified in the absence of the labels provided.

By using a set of stimuli which vary in size along an equal interval scale, it is possible to make precise predictions about the way in which the gradients of judged similarity would be expected to shift under different labelling conditions if the labels do in fact affect similarity.

Recent work by Katz (1963) has suggested that

- (a) the adding of a common label to two similar stimuli results in increased perceptual equivalence, and
- (b) adding different labels to the stimuli results in increased distinctiveness.

Her work was based on the predictions of Dollard and Miller (1950) that when labels become associated with physical objects the linguistic cues will influence the generalization gradients of the stimuli such that (a) distinctive labels will reduce the initial generalization tendencies whereas (b) similar or identical labels

will increase the amount of stimulus generalization. Katz points out that a number of studies have confirmed these predictions (e.g. Goss, 1953; Jeffrey, 1953; McAllister, 1953; Rossman and Goss, 1951; and Spiker, 1956) while others have produced negative results (e.g. Arnoult, 1953; De Riviera, 1959; Robinson, 1955). In her 1963 study she showed that when children learned to pair common, distinctive, or no labels with four highly similar nonsense shapes, the effect of the common labels was to cause the stimuli to be perceived as identical more often and to render the stimuli more difficult to discriminate in a learning task. In a later study, Katz and Zigler (1969) further demonstrated the effect, with the prediction, derived from Piaget (1951) that names may be more important for seven-year old than for nine-year old children. The subjects were required to judge the shapes as "same" or "different"; and a clear labelling effect emerged which was more pronounced for the younger children. The authors concluded that

"Older children may be relatively insensitive to the verbal predifferentiation of stimuli and only allow such labelling training to influence their perceptual judgement in those instances where few other cues for judgements are available."

Our theoretical approach differs from that of Katz et al in that we wish to examine the labelling process in terms of cognitive mechanisms as exemplified by the judgement of similarity, rather than in terms of stimulus generalization and discrimination. The approach adopted here assumes that the subject can clearly

discriminate between the stimuli used and that his judgements of similarity / dissimilarity are based, not on generalization or perceptual equivalence, but on a cognitive weighting of the two disparate sources of information present in the stimulus complex shape-plus-label. His judgement is therefore viewed as a cognitive transduction of the information rather than as a response based on confusion or inability to discriminate.

It was also proposed to use this experiment as a pilot study for possible future investigations of the role of conceptualisation or classification of the stimuli in the forming of similarity judgements.

#### EXPERIMENT 2.

The stimulus continuum chosen for this study was an equal interval series of rectangles with successively decreasing base / height ratio but constant area. The middle stimulus in the series was a square; it served as the standard against which the 8 remaining stimuli were judged. This stimulus series was considered to be potentially useful because it formed a set of stimuli which could be conceptually organised in several distinct ways:

- (a) A two-part series consisting of relatively "long thin shapes" (numbers 1, 2, 8 and 9) and "short thick shapes" (numbers 3, 4, 6 and 7). This will be referred to as the "linearity" conceptualization.
- (b) A two-part division of the stimuli into shapes lying on their sides (numbers 1 to 4) as compared with shapes 'standing on end' (numbers 6 to 9). This will be referred to as the "orientation" conceptualization.

(c) A series varying simply in base / height ratio (given that area is constant). This will be referred to as the "optimal" conceptualization.

It was then theorised that if, by use of appropriate instructional sets, we could cause subjects to categorise the stimuli in different ways, then these different categorisations may influence the judged similarity of stimuli. For example, subjects responding to the stimuli in terms of linearity conceptualization might show a displacement of points on the similarity gradient which corresponds to the boundaries between the "long thin" and the "short thick" shapes, whereas subjects responding under the "optimal" conceptualization would not be expected to show displacement.

Although these non-optimal categorisations leave room for the extension of the present experiment to investigations of the influence of various conceptualizations on similarity judgements, it was decided to limit this experiment for the present to an investigation of the influence of labels in the optimal categorisation case.

To ensure that all subjects were aware of the various ways in which the shapes could be categorised, the three conceptualizations were pointed out in the instructions to subjects, with emphasis placed on the optimal categorisation.

The aim of the experiment, then, is to determine whether or not the pairing of nonsense labels with individual stimuli will induce a shift in the similarity judgements produced when subjects compare various rectangles with a square. There are two distinct

classes of similarity judgements in this experiment:

- (i) Judgements of unlabelled rectangles against an unlabelled standard square.
- (ii) Judgements of labelled rectangles against a labelled square.

If we assume that similarity judgements are influenced by the verbal labels that are paired with the stimuli to be judged, then two predictions can be made:

- (a) Rectangles which share a common label with the standard square will show a shift towards greater judged similarity to the square in the labelled, as compared with the unlabelled condition.
- (b) Rectangles which are paired with a label which is different from that paired with the square will show a shift towards greater dissimilarity to the square in the labelled, as compared with the unlabelled condition, i.e. stimuli possessing different names should show a decrement in judged similarity.

Two distinct forms of pairings between nonsense labels and the geometric shapes were used; the counterbalanced forms 1(a) and 1(b) listed below were based on the orientation concept mentioned above while forms 2(a) and 2(b) were a counterbalanced alternating form of labelling.

STIMULI	1	2	3	4	5	6	7	8	9
Cond'n.									
1(a)	L1	L1	L1	L1	L1	L2	L2	L2	L2
1(b)	L2	L2	L2	L2	L1	L1	L1	L1	L1
2(a)	L1	L2	L1	L2	L1	L1	L2	L1	L2
2(b)	L2	L1	L2	L1	L1	L2	L1	L2	L1

Condition 1 thus allows us to examine the effects of labelling on similarity when the labels form the basis of a simple concept ("all shapes lying on side have one label, while those standing on end have the other label"). Condition 2 however is unlikely to be perceived as containing an underlying concept, since the labels alternate, along with the added difficulty that the matching pairs 1, 9; 2, 8; etc. have different labels.

A series of eight rectangles and a square were prepared and photographed to produce black shapes on a white background. The dimensions of the nine shapes were as follows: (in centimeters):

Stim. No.	1	2	3	4	5	6	7	8	9
Height	3.60	4.30	5.00	5.70	6.32	7.02	8.00	9.30	11.20
Base	11.20	9.30	8.00	7.02	6.32	5.70	5.00	4.30	3.60

The stimuli thus form a physically equal interval series in which area is constant. The standard (number 5) lies in the middle of the series and the successively longer (or taller) pairs of shapes are all clearly discriminable from the standard and from each other.

#### Apparatus:

This consisted of two tachistoscopic projectors connected to a control unit. A pair of electrically operated Alphax shutters were mounted in front of two Leitz (Wetzlar) 35mm projectors so that a standard 35mm slide could be projected for brief predetermined intervals onto the appropriate screen. The apparatus was set so that the sequence of events was as follows:

Beginning of trial: Shutter of left hand projector open to



illuminate screen for 1 second. Closure of left hand shutter followed immediately by opening of shutter of right hand projector to illuminate the screen for 1 second. (Projection of the nonsense label was always from the left hand or first projector; the geometric stimuli were always projected on the right hand screen.) This completed the presentation of the standard; a pause of approximately 4 seconds followed while the slides were replaced with the comparison stimulus then the sequence was repeated: presentation of the label for 1 second followed immediately by the rectangle which was to be judged against the standard. A pause of up to 10 seconds followed during which time subjects recorded their responses on the prepared response sheets.

The response scale used is shown in the instructions to subjects. (Category 10 was altered as it was pointed out that the phrase "completely opposite" as used in experiment 1 was in a sense inconsistent with the rest of the similarity scale.) The scale was used in conjunction with numerals to facilitate scoring of the subject's responses. Subjects were instructed to write the chosen response category number in the appropriate box for each trial.

In the initial unlabelled trials the same sequence of events was followed except that the left hand projector remained switched off to avoid a flash of white light on the screen.

Two 4' square screens were placed adjacent to each other, at eye level on a wall in front of the subjects who were seated at a table about 8 feet from the screens. To avoid dark adaption between

trials, and to allow subjects to record their own responses on the prepared data sheets a blue-tinted 50w bulb was used to illuminate the room throughout the experiment. Care was taken to avoid distortion of the projected image due to parallax by mounting the projectors at eye level and perpendicular to the screens.

The subjects were 32 college students serving to fulfil their laboratory course requirements in introductory psychology. They served as available either singly or in groups of 2 or 3. Subjects were seated at a table and presented with the following typed instructions to read:

"INSTRUCTIONS FOR PART 1. (Unlabelled shapes)

In this experiment you will be required to judge the similarity of a series of rectangles to a square. The square and rectangle all have the same area but differ in other ways. The square will be shown before each rectangle, then after the rectangle has been projected your task is to select a category from the list below to express how similar you think the rectangle is to the square.

The categories are:

1. Identical
2. Nearly identical
3. Very similar
4. Similar
5. Slightly similar
6. Slightly dissimilar
7. Dissimilar

8. Very dissimilar
9. Extremely dissimilar
10. Completely different.

Try to respond as accurately as possible and do not look at your neighbour's work. Use the sheets provided to record your responses. You need only write the number of the response category which you choose, e.g. if you judge a given rectangle to be "similar" to the square then you will write a number (4) in the appropriate space on the sheet.

PLEASE TAKE NOTICE OF THE FOLLOWING POINTS.

1. The shapes can be classified in a number of ways, e.g. we could arbitrarily divide them into

- (a) short, thick, as distinct from long, thin rectangles.
- (b) rectangles lying on their longer side, as distinct from those standing on the shorter end.
- (c) a series of rectangles with successively shorter bases (and therefore, successively longer heights, since they have constant area).

For this experiment, please try to categorise them by the 3rd method above, that is, note that they all have the same area but different ratios of base to height; that is as the base lengths successively decrease, the heights increase.

2. The standard square is always presented first; you are required to judge similarity of the rectangle to this square.

3. Before we start the experiment I will show you the whole series of slides to give you an idea of how much they differ from each other."

The 18 unlabelled trials followed after a demonstration trial on which any questions raised by subjects were answered. The standard square was always presented first (on the right hand screen) followed by the comparison stimulus (on the right hand screen). Each of the 9 comparison stimuli was presented twice on the 18 randomly ordered trials. Subjects' response sheets were then collected and fresh sheets were handed out along with the instructions for Part 2. The 36 randomly ordered labelled trials then followed and subjects were instructed to continue responding as before. The instructions for Part 2 were as follows:

"INSTRUCTIONS FOR PART 2. (Labelled shapes)

Part of this experiment is concerned with people's ability to learn nonsense syllable labels when they are paired with geometric figures.

On the left hand screen a 3-lettered nonsense syllable label will be projected at the same time as the square or rectangle appears on the right hand screen. Try to learn which syllable goes with which shape; there will be a test at the end of the experiment to see how many of the geometric figures you are able to correctly label. Please continue judging similarity of the rectangle to the square as before.

The categories are:

1. Identical
2. Nearly identical
3. Very similar
4. Similar

5. Slightly similar
6. Slightly dissimilar
7. Dissimilar
8. Very dissimilar
9. Extremely dissimilar
10. Completely different. "

The final test, in which the stimuli were presented in order 1 to 9, was included to attempt to force subjects to pay attention to the labels during the previous trials on which the similarity judgements were made. This post-test also enabled us to divide subjects into concept solvers vs. nonsolvers since in Condition 1 (easy concept) subjects either got the label-stimulus pairings completely or almost completely correct, or else they showed by their responses that they were obviously guessing.

The four counterbalanced pairings of labels with stimuli are shown above; each subject served only once in one of the 4 conditions. The labels used were selected from the list of 0% Association Value (Glaze, 1928) and were as follows: L1 : QAP  
L2 : ZIQ.

#### Results:

For each subject, the mean response category number was calculated by pooling the two responses to each comparison stimulus in Part 1 (unlabelled) and again by pooling the 4 responses to the labelled stimuli in Part 2. The mean response to the labelled stimulus was then subtracted from the mean response to the unlabelled stimulus to produce a mean category shift score for each

of the 9 stimuli that were judged by each subject. On the basis of the post test which was given to each subject at the conclusion of the experiment, subjects were divided into those who solved the concept involved (as shown by an errorless pairing of each label with each geometric shape) and those who had not learned the correct pairings during the four labelled trials on which each stimulus was presented. Approximately half of the subjects solved the relatively simple concept in Condition 1 but only one subject produced the correct pairings in Condition 2 where the obscurity of the concept involved meant that a strategy bordering on rote learning (the type VI of Shepard, Hovland & Jenkins, 1961) was the only one likely to lead to learning of the correct stimulus-label pairings.

The mean category shift scores for each subject under each condition are shown in Table 3:1.

The mean category shift scores were analysed using the discriminant analysis program outlined in the pilot taste experiment. In order to satisfy the formal statistical requirements of the method used it was considered advisable to combine concept solvers and nonsolvers into single groups and analyse only in terms of the four different conditions in which labels were differentially paired with the shapes.

The analysis yielded a Mahalanobis D-square of 36.57 which is distributed as Chi-squared with 27 degrees of freedom. This falls far short of statistical significance. However, when the mean Bayesian a posteriori probabilities were calculated by pooling the probabilities given by the 8 subjects in each group a fairly

TABLE 3:1    MEAN CATEGORY SHIFTS BETWEEN JUDGEMENT OF UNLABELLED  
AND JUDGEMENT OF LABELLED SHAPES FOR EACH OF THE 8 SUBJECTS IN  
THE 4 EXPERIMENTAL CONDITIONS.

Stimulus	1	2	3	4	5	6	7	8	9
S1	1.00	1.25	2.25	1.50	0.75	-2.50	-0.25	-0.25	0.00
S2	0.50	-1.00	-1.00	0.00	0.25	-0.50	-1.00	-1.00	-0.25
S3	-1.50	0.25	1.00	0.25	1.00	0.25	0.00	-0.25	-0.75
S4	0.25	2.00	0.75	1.50	0.75	-0.50	3.25	1.50	-0.50
S5	1.50	1.00	0.75	1.50	-1.50	-0.50	-0.50	-0.25	-0.25
S6	-1.25	-1.75	-0.25	1.75	1.25	0.75	1.25	-1.25	-0.75
S7	1.50	0.00	0.00	0.50	-0.50	-0.25	1.00	0.75	0.25
S8	0.00	0.00	0.25	0.25	0.50	0.25	0.00	-1.75	-1.00
S9	0.00	1.75	-0.25	0.75	0.00	-1.00	1.00	-0.25	0.00
S10	1.25	0.50	1.75	1.50	0.50	0.75	2.25	0.00	1.50
S11	0.00	0.00	0.75	-1.50	-0.25	-4.00	2.50	1.50	1.50
S12	0.50	0.50	1.75	1.50	0.00	-0.25	1.00	-1.00	0.25
S13	-1.25	-0.75	-1.00	2.00	-0.50	-1.50	-1.25	-0.25	0.00
S14	0.25	0.25	0.75	1.25	-0.50	-0.25	1.00	-0.50	-1.00
S15	0.00	-0.25	2.75	1.00	1.50	4.25	4.00	0.00	-0.75
S16	1.00	3.25	-0.25	0.00	0.50	0.00	-0.50	0.75	0.75

Table 3:1.

Stimulus	1	2	3	4	5	6	7	8	9
S17	0.50	0.00	2.75	1.75	0.00	0.50	2.25	0.50	0.00
S18	0.50	1.75	1.25	-0.25	-0.50	2.00	0.25	-0.50	0.50
S19	1.50	0.25	0.00	1.75	0.50	2.25	1.50	1.00	0.00
S20	1.00	1.25	1.00	1.00	0.25	0.00	0.00	1.75	0.00
S21	1.25	0.00	1.00	1.75	0.50	0.75	-0.75	-0.25	0.00
S22	0.00	0.25	2.00	0.50	0.00	-0.25	1.25	0.50	0.50
S23	0.00	-1.50	2.25	2.50	0.50	-0.25	1.25	-0.75	0.00
S24	1.50	0.00	0.00	1.50	0.25	-1.00	-0.25	0.00	-0.25
S25	0.50	-0.25	0.50	1.25	1.00	0.50	1.50	1.00	0.75
S26	0.00	-0.75	-0.50	2.00	0.00	1.25	0.25	0.00	0.00
S27	0.00	0.25	0.75	0.25	0.50	-1.75	0.25	-0.50	-0.75
S28	1.75	2.00	2.50	1.25	0.00	0.25	0.25	1.25	0.00
S29	-0.50	-1.00	0.50	0.25	0.25	-0.25	-0.50	-1.25	-0.50
S30	1.00	1.00	0.00	0.50	0.00	1.25	0.75	0.50	0.50
S31	-0.25	-1.00	2.50	0.25	0.00	0.00	-1.00	1.25	-1.00
S32	0.50	0.50	3.50	3.25	1.00	1.25	0.00	-0.25	-0.25



clear trend in the expected direction emerged. These probabilities are plotted as histograms in figure 3:1. The most clear-cut separations appear in the two counterbalanced groups in Condition 1, where the mean probability associated with each appropriate group shows a relatively large separation from the three remaining groups in each case. The results for the two counterbalanced groups in Condition 2 show a less clear-cut separation. This could be taken as slight evidence that the subjects in Condition 2 were faced with a conceptually more difficult task; hence they performed less satisfactorily on it and thus provided less statistical basis for separating out differences in response profile in their data.

The mean judgements to each labelled and unlabelled stimulus for the 8 subjects in each of the 4 experimental groups are plotted in figure 3:2. Note that in each case the judgements made to each stimulus (as compared against the standard) tend to show a general enhancement of similarity for each shape regardless of which label was paired with the shape. There are a number of possible explanations for this apparent shift in judgements between the labelled and the unlabelled conditions. The fact that the judgements between labelled stimuli were always made following the judgements of the unlabelled shapes suggests that there may be a consistent tendency for subjects to redistribute their responses on the scale with successive exposures to the stimuli over time. For example, if on repeated exposure to the stimuli, subjects tended to see greater similarity between the square and the rectangles, we would obtain the effect shown in figure 3:2. However, this question can be

MEAN BAYESIAN A POSTERIORI PROBABILITY FOR  
MEMBERSHIP IN EACH A PRIORI (TREATMENT) GROUP.

FIG 3:1

MEAN A POSTERIORI GROUP MEMBERSHIPS  
PRODUCED BY DISCRIMINANT FUNCTION  
ANALYSIS OF DATA FROM EXPERIMENT 2.

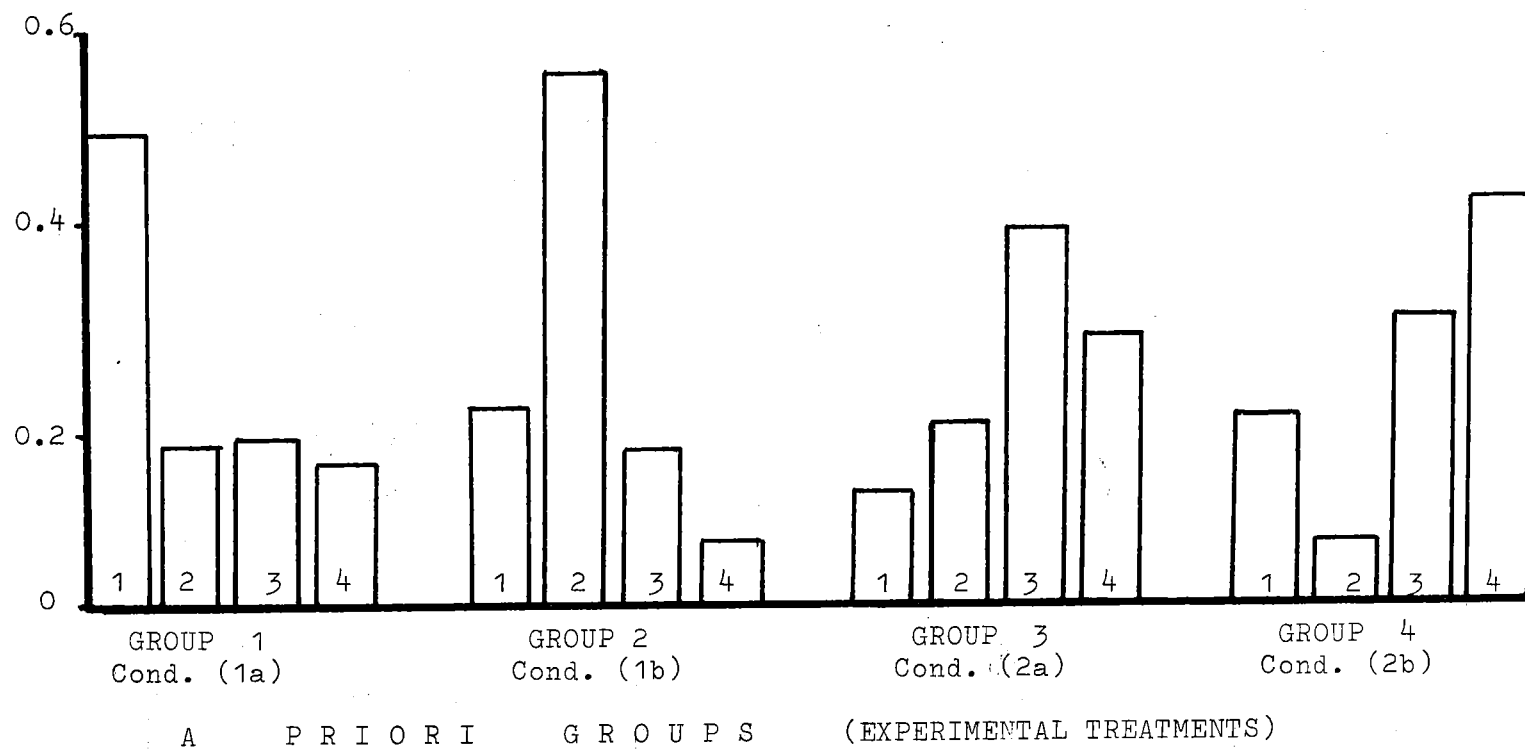


FIG. 3:2 EXPERIMENT 2 - EFFECTS OF LABELLING ON  
JUDGEMENT OF SIMILARITY BETWEEN LABELLED  
SHAPES. - CONDITION 1(a)

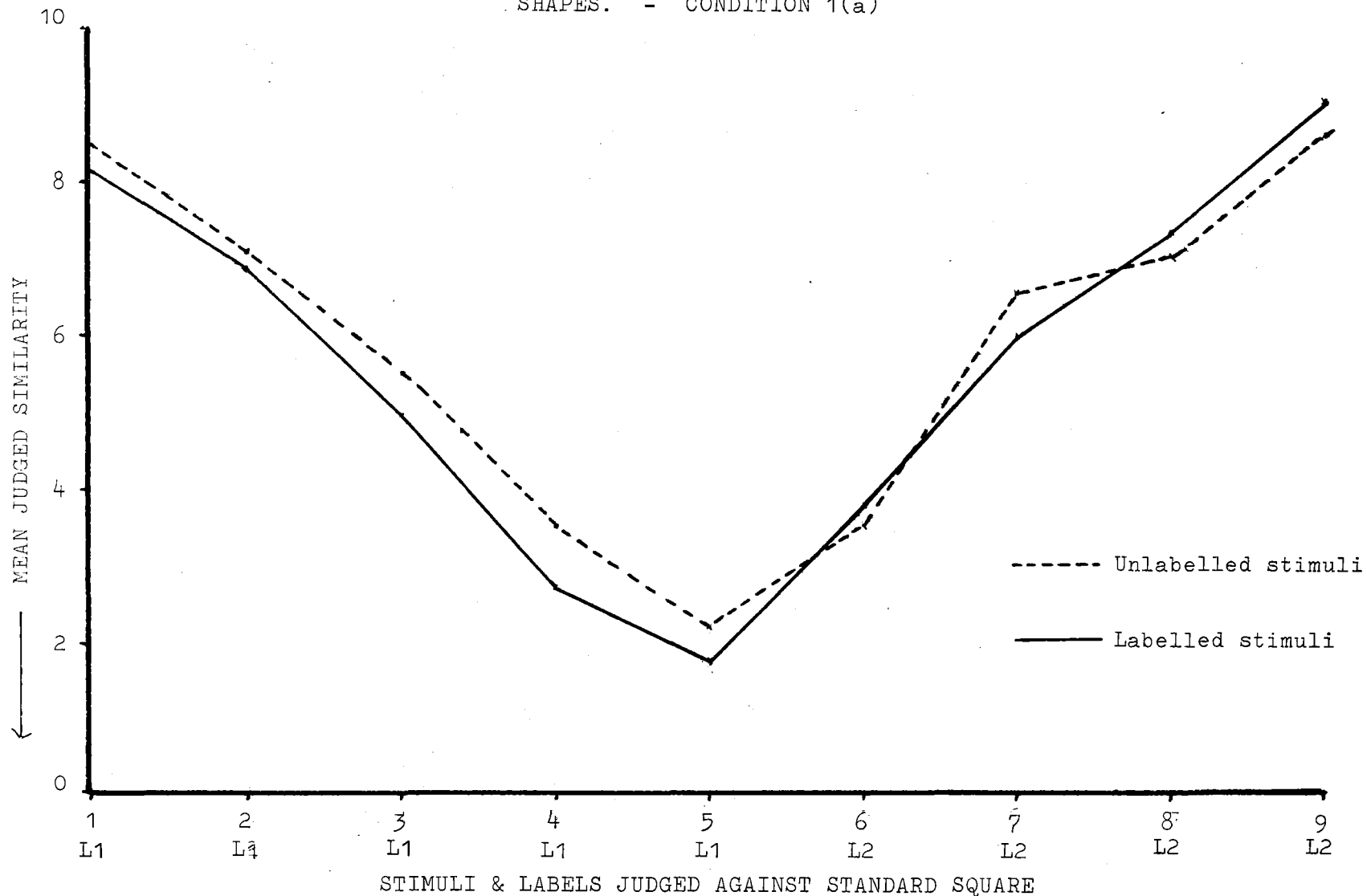


FIG. 3:2 EXPERIMENT 2 - EFFECTS OF LABELLING ON  
JUDGEMENT OF SIMILARITY BETWEEN LABELLED  
SHAPES. - CONDITION 1(b)

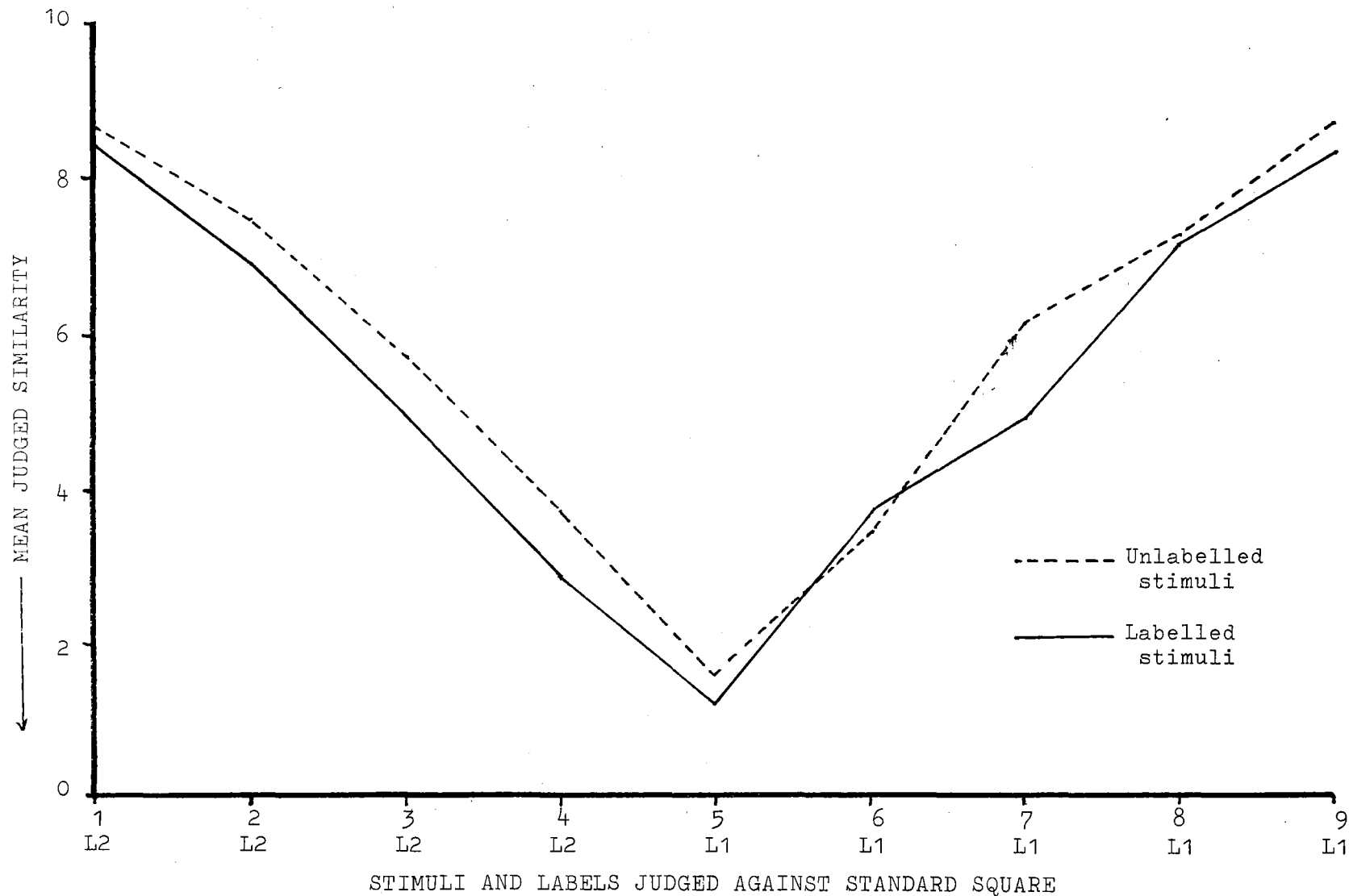


FIG 3:2 EXPERIMENT 2 - EFFECTS OF LABELLING ON  
JUDGEMENT OF SIMILARITY BETWEEN LABELLED  
SHAPES. - CONDITION 2(a)

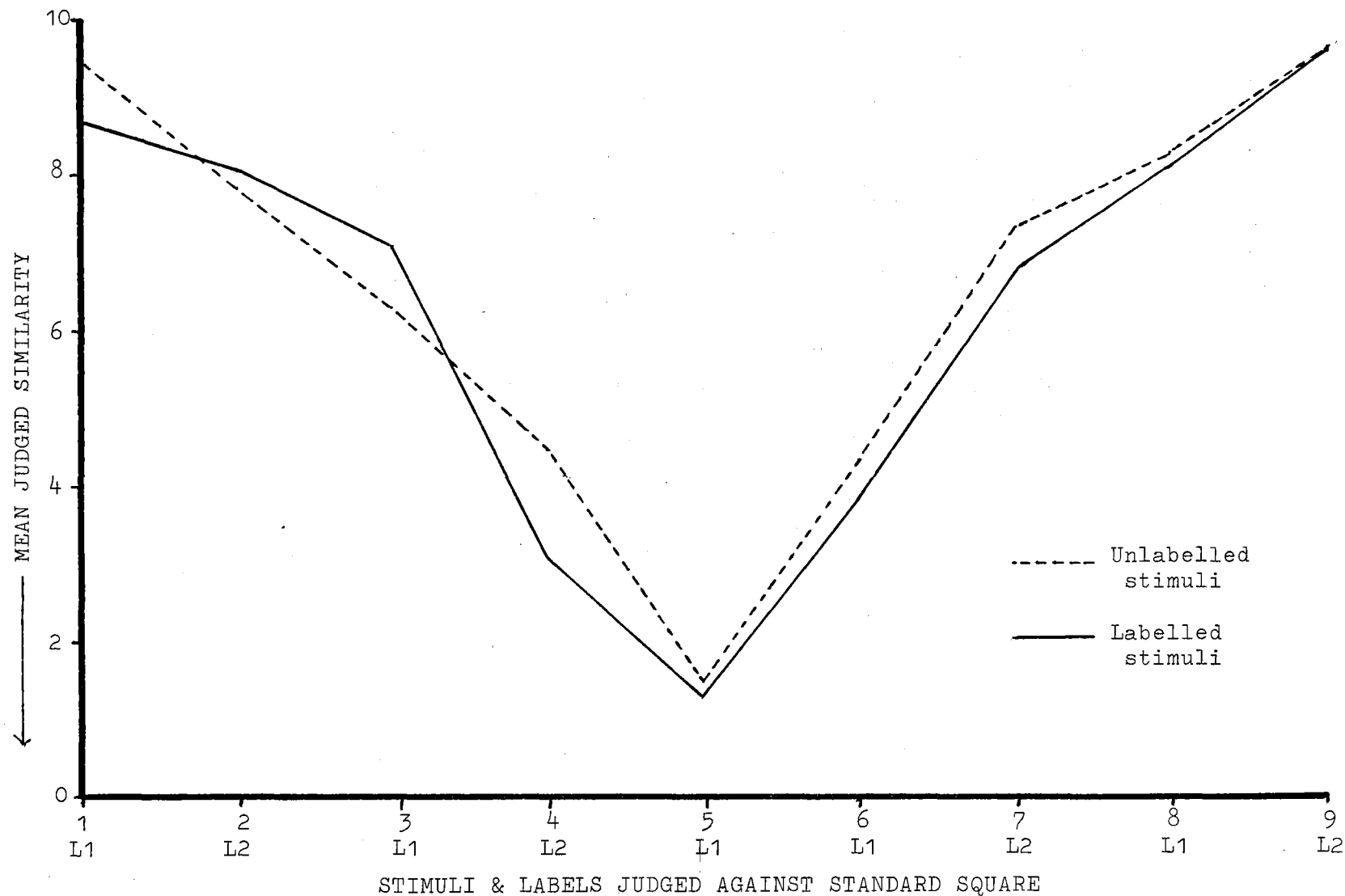
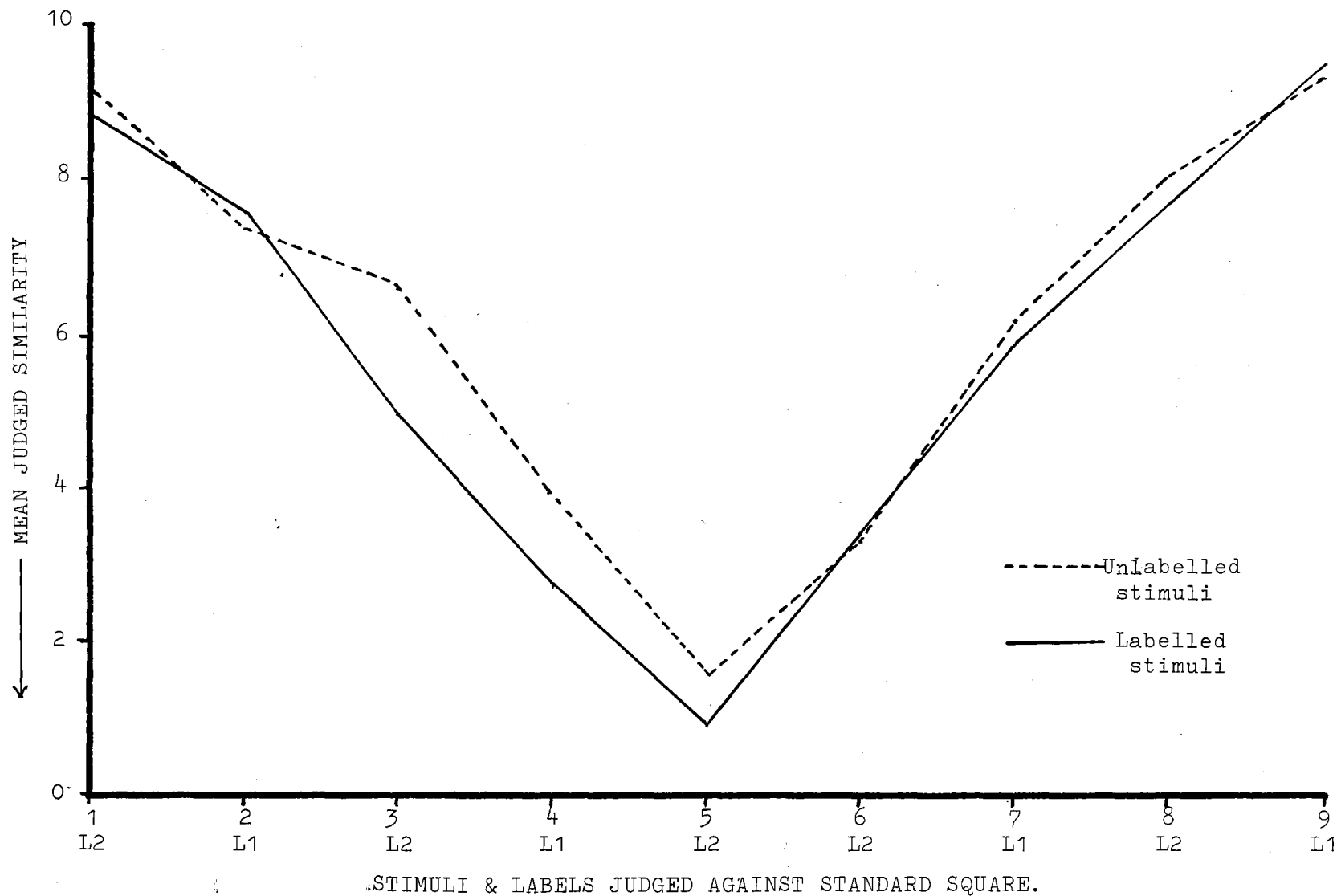


FIG. 3:2 EXPERIMENT 2 - EFFECTS OF LABELLING ON  
JUDGEMENT OF SIMILARITY BETWEEN LABELLED  
SHAPES. - CONDITION 2(b)



answered by referring back to the original data. If subjects tend to use successively lower categories on the ten point scale with repeated exposure to the stimuli, then this trend should show up within the 4 repeated trials that were given for each stimulus pair under the labelled condition. If we take the 1st and 4th presentation of each labelled stimulus pair in turn, we can tally the frequencies with which the response given to the first presentation of the stimulus exceeds in scale value the response given to the same labelled stimulus on its 4th presentation.

For the 4 experimental conditions, these frequencies were as follows:

		Frequency	
Condition 1(a)		19 out of 72	or 27.7%
"	1(b)	11 out of 72	or 15.3%
"	2(a)	23 out of 72	or 31.9%
"	2(b)	26 out of 72	or 36.1%

There does appear to be a slight tendency for judgements to shift over repeated trials however the table shows that the shift is in the direction of greater judged dissimilarity over successive trials whereas the graphs in Figure 3:2 show a shift in the opposite direction. We can thus rule out the possibility that a progressive shift, over trials, towards greater judged similarity accounts for the enhanced similarity shown for judgements between labelled as distinct from unlabelled stimuli.

This lends further support to the interpretation that the shift in judgements can be attributed to a labelling effect. Since,

in the majority of cases, the labelled stimuli are judged to be more similar to the labelled square than is the case with the unlabelled shapes, it seems reasonable to suggest that the effect of labelling is to produce a general enhancement of similarity regardless of the "meaning" that we have attempted to attach to the nonsense labels. That is, we have again demonstrated a phenomenon that is similar to the effect reported in the pilot study, where nonsense labels seem to act as a second stimulus dimension and reduce the judged difference between the stimuli, by overwriting any real-word categorisation that subjects may have assigned to the stimuli in the absence of nonsense labels. That is, from the subject's point of view, the labels ZIQ and QAP are handled as equivalent, (but secondary) cues in the judgement situation.

As a statistical test of the hypothesis that the nonsense labels have enhanced the judged similarity of the rectangles to the standard square, a tally was made of the frequencies with which, for each subject's judgement of each stimulus pair, the mean unlabelled judgement exceeded in scale value the mean labelled judgement. The number of tied pairs was subtracted to give 162 out of 240 or 67.5% of the differences in the predicted direction. The Binomial test gave a z-value of 5.36, which is significant beyond the 0.001 level. There is thus considerable support for the hypothesis that the labels function as a kind of second dimension, in which the two labels ZIQ and QAP are functionally equivalent, and by virtue of the fact that they apparently add a perceptually identical cue to all of the shapes, they enhance the



perceived similarity of the labelled shapes. Since there was minimal delay between presentation of the standard, and presentation of the comparison stimulus, the observed shift in judgements cannot be attributed to the effects of labels on trace decay.

A separate analysis of the data from Condition 1 was carried out to investigate the possibility that those subjects who had solved the "concept" may have responded differently from those who were unable to remember which labels had been paired with each shape. No systematic tendency emerged however; perhaps for some of the following reasons.

1. Our criterion for "concept" solution was not particularly stringent. Many of the subjects who solved the "concept" may have done so in the final trials of the experiment only, so that the majority of their responses were made in ignorance of the system under which the shapes and labels had been paired.
2. By pairing labels with shapes in this way we are attempting to force subjects to reorganise their conceptualization of the underlying dimensionality of the stimuli. In this sense we are pitting a few trials of experimental manipulation against a lifetime of learned conceptual organising responses. It is not surprising, therefore, that the nonsense syllables did not come to acquire meaning. The manipulation may have been more effective if we had used labels like 'horizontal' and 'vertical' to attempt to force subjects to recategorise the shapes.
3. The way in which subjects categorise the stimuli may not be related to the outcome of the similarity judgement task.

Gregson (in preparation) has pointed out that it is possible to take a theoretical position which states that similarity as such is a purely subjective event which exists entirely within the organism in much the same way as the perception of aesthetic appeal is considered to operate. It can be argued however, that all perception is in this sense subjective, in that the classification of the outside world into concepts, classes or stimulus relations exists only within the observer. The relationship between similarity perception and the perception of conceptual relations is considered to be theoretically significant by this writer, in that although the methodological and theoretical approaches to both are considered to be distinct, the cognitive mechanisms underlying the two processes may be closely related. Posner, Goldsmith and Welton (1967) have attempted to develop a psychophysics of form-similarity which is appropriate to the study of concept learning. A form of cognitive classificatory behaviour which has been termed schema theory by Evans (1967) can be considered to be related to concept learning in the more traditional sense. Rankin, Markley and Evans (1970) have shown that pretraining of subjects in the identification of schematic concepts significantly increases the judged similarity of the pairs of items comprising the concept. These findings appear to implicate similarity processing in conceptual behaviour. Brown and Dansereau (1970), however, report that somewhat different features of the stimuli are used in making classification responses and similarity judgements. This however, does not necessarily mean that the cognitive processes used in similarity assessment and

conceptual behaviour are different. However, the postulate that related cognitive mechanisms are used in the two processes would be strengthened if it could be shown that the verbal processes which are considered to play a significant role in concept formation could also be shown to play the same part in the perception of similarities. It was decided however, to limit this research to the investigation of verbal processes in similarity judgement for the present.

## CHAPTER IV

### SOME EFFECTS OF QUALITATIVE LABELS ON

#### TASTE SIMILARITY JUDGEMENTS

The experiments described in Chapters 2 and 3 made use of nonsense syllables as labels for the stimuli. It was pointed out that the apparent failure of nonsense labels to acquire meaning as a consequence of being paired with stimuli may be due to the fact that subjects already possess highly learned or overlearned labels for the familiar kinds of stimuli used in these experiments, and that these pre-experimental labels are difficult to overwrite during the few trials which are available in an experimental setting. An alternative approach is to use potentially unfamiliar stimuli and label them in a variety of ways using real word labels. If labelling is important in the categorisation or classification of stimuli, and this categorising process influences the judged similarity between the stimuli, then the judgements produced by groups of subjects under various labelling conditions will presumably vary systematically from group to group. It is assumed that when a subject is confronted with an unfamiliar stimulus he will categorise it in terms of other more familiar stimuli. However, we need to determine whether or not this categorisation process can be influenced by using experimenter-provided labels.

Recently, a new flavouring material has become available

which is rarely found as an isolated taste in everyday life and which has complex qualitative characteristics. The substance is a nucleotide called Disodium 5' Guanylate (abbreviated D5'G) which is available in highly purified form; it is a white powder which is both readily soluble and fairly stable in aqueous solution.

As a prerequisite to the experiment described in this chapter, it was necessary to have some knowledge of the qualitative gustatory characteristics of this substance. An experiment was conducted jointly by Gregson and the author to determine these characteristics. The experiment is described elsewhere (Gregson and Simmonds, 1971).

In the concluding pages of Chapter 2, the role of memory in similarity judgement tasks was discussed in some detail. It was suggested that a temporal decay in memory traces might account for much of the variability in the responses produced by subjects in the pilot study, but that the pairing of labels with stimuli may help to counteract the effects of trace decay. It was decided in the experiment which follows, to return to the experimental approach used in the pilot study, with gustation again as the sense modality to be investigated.

### EXPERIMENT 3

The aim of this experiment is to investigate the effects of verbal labels (which vary in the degree of precision or adequacy with which they conceptualise the taste qualities present in a set of tastes) on the judged similarity between the tastes. It is postulated that the verbal input which can be made to

accompany the taste solutions may influence the processing of the information contained in the input, when complex cognitive manipulation of this information is required, such as is presumed to take place during the process of similarity judgement.

The pilot study showed that there is an appreciable error or noise level in the responses produced when subjects judge the similarity between pairs of tastes. This does not appear to be entirely attributable to confusion or failure to discriminate between taste sensations at the sensory input stage of the process, since the four tastes used in Condition 1 of the experiment were clearly discriminable (the four primary taste qualities were used and all were well above threshold concentration). In keeping with the theory of trace decay outlined by Brown (1958), it is hypothesised that part of this noise level may be attributable to decay in, or confusion between the memory traces involved when a taste which is present in the mouth is judged against a set of standards which are carried in the memory. When labels are paired with the stimuli they are considered to slow down the decay of the memory trace and hence facilitate the judgement task. Presumably, the more information carried by the labels, the more accurately will the memory trace be retained as shown by the degree of discrimination between pairs of tastes. In particular, judgements made between physically identical pairs of stimuli should show a decrease in error variance as we increase the degree to which various qualitative aspects of the taste stimuli are

captured by the labels. As well as the effect of labelling on judgements between physically identical tastes a number of other labelling effects will also be investigated. These include:

- (a) the effect of slightly inappropriate or erroneous labels on judgements between tastes
- (b) the effect of identical labels on tastes of varying degrees of physical dissimilarity
- (c) the effect of labels which vary in the number of taste qualities which they describe in the solutions used in the experiment.

A qualitative rating scale will also be used as a post-experimental check on the qualitative characteristics of the taste stimuli used. A copy of this rating scale is shown in Appendix I.

#### PROCEDURE

Four master solutions were prepared in the following concentrations, using deionised water:

3% NaCl	(0.03 gm/ml)
0.5% citric acid	(0.005 gm/ml)
0.5% sodium benzoate	(0.005 gm/ml)
0.04% D5'G	(0.0004 gm/ml)

Two mixtures were prepared from the NaCl and citric acid master solutions to give a total of 6 taste stimuli:

Standard 1: 3 volumes of NaCl to 1 volume of acid

Standard 2: 1 volume of NaCl to 3 volumes of acid

Standard 3: Sodium benzoate (0.5%)

Standard 4: D5'G (0.04%)

Stimulus 5: NaCl (3.0%)

Stimulus 6: Citric acid (0.5%)

Stimuli 1 to 4 served as the four standards against which the total set of 6 comparison stimuli were judged.

The choice of stimulus concentrations was largely arbitrary, but nevertheless was influenced by the following considerations: Stimulus 1 and stimulus 2, the two mixtures, were made from master solutions of NaCl and citric acid. The concentrations used were the same as those used in the original pilot study; these appear in Gregson (1968) and were in turn based on a study by Beebe-Center and Waddell (1948). They are considered to be approximately of equal subjective intensity. However, when mixed in equal proportions they do not apparently produce an equally salty and sour mixture and in fact even at the 3 to 1, and 1 to 3 mixtures used here there seems to be a tendency for the salty component to dominate (as shown by the subjective rating scale responses given by subjects from the unlabelled conditions). The intention, however, was merely to create a fairly similar pair of stimuli for which several labels were more or less appropriate or relevant (e.g. "salty", "sour", "salty/sour"), in the sense that the labelled component was at least detectably present, if not the dominant component of the



solution. (Again, the rating scale responses gathered from subjects under the "no labels" conditions suggest that both components could be detected in both mixtures.)

Stimulus 3, sodium benzoate, is a peculiar substance in that it elicits all four primary taste sensations at the 0.5% concentration used here, though it is predominantly sweet, and slightly less predominantly salty at this intensity. (See Peryam, 1960; also Amerine, Pangborn and Roessler, 1965, p.115; and Gregson, 1969).

Stimulus 4, Disodium 5'Guanylate, also elicits the 4 primary taste sensations, although it is not certain as yet, how much of this response pattern is attributable to noise (Gregson and Simmonds, 1971). The concentration used here is the same as the strongest concentration used in the Gregson and Simmonds study, although this concentration is considerably nearer the threshold than the concentrations used for any of the other 5 stimuli in this experiment. Although this relatively low concentration may have caused some subjects to fail to detect the presence of the D5'G taste, it was considered desirable to have at least one low-intensity solution, since this may provide more latitude for a "labelling" effect than would be possible with a more clearly supra-threshold sweet-tasting solution.

In order to evaluate the psychophysical method used in the original pilot study, and again in this study, and also to gain information about the nature of the taste memory trace when

standards are carried in the memory, it was decided to use 20 subjects as a comparison group using a form of the more traditional method of paired comparisons.

#### THE PAIRED COMPARISONS GROUP

Ten permutations of a set of  $2^4$  randomly ordered pairs of stimuli were prepared; two subjects responded to each of these 10 different orders of presentation. Although the method of paired comparisons is rather cumbersome and does not generate much data in this context (one response for every two tastes, plus rinses between tastes) every effort was made to keep the method as closely parallel to the "memorised standards method" as possible. Thus to equate the amount of pre-judgemental experience of the 4 standards in both methods, the subjects tested under the paired comparisons method were given the same block of initial discrimination learning trials, in which the names "Standard 1", "Standard 2", "Standard 3", and "Standard 4" were associated with the 4 standards to a criterion of at least one errorless identification of the set of 4 standards. To counteract for any tendency towards asymmetry of judgement of pairs (i.e. A cf B may not produce the same judgement as B cf A), within certain limits which are outlined below, the standard was presented once as the first taste in each pair, and once as the second taste in the pair. The limitations mentioned are that, in order to cut down on the number of tastings and rinses made by each subject, the self-

comparison pairs (Stimulus 1 with 1; 2 with 2, etc.) were presented only once in the series. Also since stimuli 5 and 6 are not used as standards in the "memorised standards" method with the result that there is no counterbalancing for asymmetry there, stimuli 5 and 6 were never presented as the first of a pair in the method of paired comparisons used here.

The total set of paired comparisons made by each subject then (shown here in nonrandomised order) was as follows:

1:1	2:1	3:1	4:1	1:4	2:4	3:4	4:4
1:2	2:2	3:2	4:2	1:5	2:5	3:5	4:5
1:3	2:3	3:3	4:3	1:6	2:6	3:6	4:6

The instructions to subjects used in the method of paired comparisons were as follows:

#### "INSTRUCTIONS FOR METHOD OF PAIRED COMPARISONS"

As part of this research programme, it is necessary to determine the similarity between a number of different tastes. These tastes will be presented to you in pairs; you are required to judge the similarity of the second taste in each pair to the first taste in that pair. Please communicate your judgement by pressing the appropriate key on your right. For example, if you decide that the second taste of a pair is identical to the first taste then you will press the key labelled "identical". If you think the two tastes are completely different, then you will press the key opposite the label "completely different", and so on.

Please try to make your judgements as accurate as possible, and follow the sequence of events as outlined below:

GREEN LIGHT: Rinse with water, rest and wait.

RED LIGHT: Pick up the sample and prepare to taste.

YELLOW LIGHT: Take the sample into the mouth, roll it around the tongue, and note its precise taste.

GREEN LIGHT: Spit out in the cuspidor, rinse with water, and wait.

RED LIGHT: Pick up the second sample of the pair and prepare to taste.

YELLOW LIGHT: Taste the sample, and note its similarity to the first taste of the pair.

GREEN LIGHT: Hold the solution in your mouth while you make the key-press (I will press the buzzer when I want you to respond), then spit out, rinse, and rest until we are ready to go on to the next pair.

The tastes are all perfectly pure and harmless substances.

#### ADDITIONAL NOTE

Before we go on with the main part of the experiment as outlined above, it is necessary that you become familiar with four of the tastes that I will be presenting to you. These tastes will be presented to you in beakers labelled Standard 1, Standard 2, Standard 3, and Standard 4. It is essential that you learn which name belong to each taste before we go on with the main part of the experiment. (To test you ability to identify and name each taste, I will sometimes present them to you in unlabelled beakers and in different orders.) "

The tastes were presented approximately at the rate of one

rinse plus taste per minute; the sequence of events is shown in the instructions. Each subject also completed one form of the qualitative rating scale for each of the four standards (i.e. four forms per subject). The information obtained from these ratings provided a pool of suitable labels to be used in the "memorised standards" part of the experiment. To facilitate choice of the dominant taste sensation present in each standard, eight of the subjects were instructed to tick only one descriptive category on each form for each standard.

#### THE "MEMORISED STANDARDS" GROUP

The procedure for this group was basically the same as that used in the original pilot study. As with the paired comparisons group, subjects first learned to identify the four standards by name to a criterion of at least one errorless trial. The actual names used to label each taste were then pinned inside the subject's cubicle at eye level in front of him; the labels used are shown for all four experimental conditions shown in Table 4:1.

TABLE 4:1

	"No Labels"	"Contradicting Labels"	"Partial Labels"	"Complete Labels"
STANDARD	CONDITION 1	CONDITION 2	CONDITION 3	CONDITION 4
1. (NaCl + Acid)	"Standard 1"	"Salty Standard 1"	"Salty Standard 1"	"Salty/Sour Standard 1"
2. (Acid + NaCl)	"Standard 2"	"Sour Standard 2"	"Sour Standard 2"	"Sour/Salty Standard 2"
3. (Sod. Benz.)	"Standard 3"	"Sweet Standard 3"	"Sweet Standard 3"	"Sweet/Bitter Standard 3"
4. (D5'G)	"Standard 4"	"Sweet Standard 4"	"Sweet Standard 4"	"Bitter Standard 4"
COMPARISON STIMULUS	CONDITION 1	CONDITION 2	CONDITION 3	CONDITION 4
1. (NaCl + Acid)	-	"Salty/Sour"	"Salty"	"Salty/Sour"
2. (Acid + NaCl)	-	"Sour/Salty"	"Sour"	"Sour/Salty"
3. (Sod. Benz.)	-	"Sweet/Bitter"	"Sweet"	"Sweet/Bitter"
4. (D5'G)	-	"Bitter"	"Sweet"	"Bitter"
5. (NaCl)	-	"Pure Salty"	"Salty"	"Pure Salty"
6. (Acid)	-	"Pure Sour"	"Sour"	"Pure Sour"

In all conditions except for Condition 1, the comparison stimuli were also labelled; sometimes with the corresponding tastes bearing the same labels as the standards (Conditions 3 and 4), sometimes with different labels (Condition 2) or with no labels (Condition 1).

Condition 1 - the "no labels" condition - serves as a control on the labelled conditions, and also allows a direct comparison between the "memorised" standards method and the method of paired comparisons. Condition 2 will be referred to as the "contradicting labels" condition as the standards and the comparison stimuli are labelled in slightly different ways. Note that the D5'G is labelled as "Sweet" when tasted as a standard, but is labelled "Bitter" when it appears in the comparison series.

In Condition 3, the labels only partially capture the total taste sensations that are present in the solutions - this is referred to as the "partial labels" condition.

The labels used in Condition 4 more completely describe each taste, and, unlike Condition 3, no two tastes within the six comparison series tastes share the same label. Condition 4 will thus be referred to as the "complete labels" condition.

In all 4 conditions, each subject judged each of the six comparison stimuli against the four memorised standards a total of 4 times in random order. Five different permutations of the  $6 \times 4 = 24$  comparison stimuli were prepared and two subjects served under each permutation, giving 10 subjects in each of the four labelling conditions.

The "memorised standards" method has been summarised in Chapter 2. The essential features of the method are that, after the initial discrimination training with the standards, they are not experienced as such again, although they occur in physically identical form in the comparison series. (Subject is not informed

that these tastes are identical to the standards). The subject's task, on tasting a given comparison stimulus, is to judge the similarity of this taste, while he holds it in his mouth, against his memory trace of the four standards as initially experienced. The four judgements are made consecutively, always responding in the order of comparison of the stimulus with Standard 1 followed by 2, 3, and 4. The tastes were presented at the rate of approximately one per minute; there was a rinse between each taste and the 4 responses to each taste were made during the minute. Other details can be found in the instructions to subjects, which are given below.

#### "INSTRUCTIONS FOR "MEMORISED STANDARDS" METHOD

In this experiment you will be required to judge the similarity between various tastes by responding on the bank of keys on your right. You will first be presented with a set of four tastes called standards and your task is to judge the similarity of a number of tastes to each standard in turn. To make the task as simple as possible for you I will present the tastes in labelled beakers. It will be necessary for you to first learn which name goes with each one of the standards, which are clearly labelled Standard 1, Standard 2, Standard 3, and Standard 4. The same name will always be paired with the same taste, and it is essential that you learn the name of each one of these tastes before we proceed further with the experiment. (I will sometimes present the standard without giving you its name



to see if you can correctly identify it.)

After you have learned the names for each of the standards 1 to 4, I will then present another set of tastes, and you are to judge how similar each taste is to each of the four standards in turn. That is, you will make four key-presses for every taste in turn; the first keypress expressing the similarity of the present taste to Standard 1, the second press representing the similarity of this taste to Standard 2, and so on for Standards 3 and 4. Please hold the taste in your mouth while you are making your responses, and follow the sequence of events as listed in the table below. The tastes are all perfectly pure and harmless substances.

#### SEQUENCE OF EVENTS FOR SIMILARITY JUDGEMENTS

GREEN LIGHT: Rest

YELLOW LIGHT: Pick up beaker, read the label, and prepare to taste.

RED LIGHT: Take full contents of beaker into mouth, roll it around the tongue, and note its precise taste as compared with the other tastes that you have experienced so far in this experiment.

GREEN LIGHT: Make judgements by pressing the appropriate keys.  
I will press the buzzer four times:

1st buzz: Judge similarity of taste to Standard 1

2nd buzz: Judge similarity of taste to Standard 2

3rd buzz: Judge similarity of taste to Standard 3

4th buzz: Judge similarity of taste to Standard 4

Spit out in cuspidor, rinse with water, then rest and wait for the next stimulus."

## RESULTS

The results were transposed directly from Esterline Angus charts to 5 4 x 6 matrix charts. The first 4 rows of each 6 x 4 matrix were condensed into a half matrix by combining symmetrically opposite pairs of judgements to give five 18 cell tables of the following form:

		Standards			
		1	2	3	4
Stimuli	1	1:1	1:2	1:3	1:4
	2		2:2	2:3	2:4
	3			3:3	3:4
	4				4:4
	5	1:5	2:5	3:5	4:5
	6	1:6	2:6	3:6	4:6

In order to determine whether or not the judgements between stimuli were influenced by the labels paired with the stimuli, four discriminant function analyses were carried out to compare the judgements made between a given standard and the six comparison stimuli under the 4 labelling conditions. Thus, for example, analysis 1 compared the judgements made by the 10 subjects in each of the 4 labelling groups when Standard 1 was compared with each of the 6 comparison stimuli. The data from the paired comparisons control group was not included in these

analyses as any differences in the data produced by the method used may have spuriously increased the value of the Mahalanobis D-squares produced by the analyses. The results of the four analyses are shown below:

	<u>Analysis number:</u>	<u>Mahalanobis D-squared</u>
1.	(salty/sour standard)	102.56
2.	(sour/salty standard)	64.56
3.	(sodium Benzoate standard)	43.95
4.	D5'G standard)	55.82

The Mahalanobis D-squared is distributed as chi-squared with 18 degrees of freedom. All 4 analyses were significant beyond the 0.005 level.

Table 4:2 shows the frequencies with which the 10 subjects in each experimental group were in fact assigned to each group on the basis of the largest discriminant function calculated from their particular sets of similarity judgement scores. (The figures in parenthesis give the mean Bayesian probability associated with the largest discriminant function for the n subjects that were assigned a posteriori to each group.)

TABLE 4:2ANALYSIS 1

(Salty/Sour Standard)

Experimental Group	No. of subjects assigned to each group		
1. (Contradicting labels) (abbreviated K.L.)	Group 1	8	(0.8542)
	Group 2	1	(0.8208)
	Group 3	1	(0.8467)
	Group 4	0	-
2. (no labels) (N.L.)	Group 1	1	(0.9301)
	Group 2	7	(0.7975)
	Group 3	1	(0.4218)
	Group 4	1	(0.7276)
3. (Partial labels) (P.L.)	Group 1	2	(0.6261)
	Group 2	0	-
	Group 3	8	(0.7959)
	Group 4	0	-
4. (Complete labels) (C.L.)	Group 1	0	-
	Group 2	0	-
	Group 3	0	-
	Group 4	10	(0.8147)

TABLE 4:2ANALYSIS 2

(Sour/Salty Standard)

Experimental Group	No. of subjects assigned to each group		
1. (Contradicting labels)	Group 1	6	(0.6057)
	Group 2	2	(0.6723)
	Group 3	1	(0.6405)
	Group 4	1	(0.4737)
2. (No labels)	Group 1	3	(0.5708)
	Group 2	5	(0.7710)
	Group 3	0	-
	Group 4	2	(0.6249)
3. (Partial labels)	Group 1	4	(0.4421)
	Group 2	0	-
	Group 3	5	(0.8335)
	Group 4	1	(0.4091)
4. (Complete labels)	Group 1	1	(0.5289)
	Group 2	0	-
	Group 3	0	-
	Group 4	9	(0.6775)

TABLE 4:2ANALYSIS 3

(Sodium Benzoate Standard)

Experimental Group	No. of subjects assigned to each group		
1. (Contradicting labels)	Group 1	8	(0.6302)
	Group 2	0	-
	Group 3	2	(0.6012)
	Group 4	0	-
2. (No labels)	Group 1	1	(0.4814)
	Group 2	4	(0.7357)
	Group 3	1	(0.7525)
	Group 4	4	(0.5920)
3. (Partial labels)	Group 1	2	(0.5036)
	Group 2	1	(0.5128)
	Group 3	6	(0.5725)
	Group 4	1	(0.5298)
4. (Complete labels)	Group 1	1	(0.3937)
	Group 2	2	(0.5599)
	Group 3	2	(0.3546)
	Group 4	5	(0.6308)

TABLE 4:2ANALYSIS 4

(D5'G Standard)

Experimental Group	No. of subjects assigned to each group and mean probability associated with the largest discriminant function.		
1. (Contradicting labels)	Group 1	6	(0.9740)
	Group 2	3	(0.3699)
	Group 3	1	(0.3152)
	Group 4	0	-
2. (No labels)	Group 1	0	-
	Group 2	3	(0.8536)
	Group 3	6	(0.4661)
	Group 4	1	(0.5380)
3. (Partial labels)	Group 1	2	(0.5694)
	Group 2	1	(0.4891)
	Group 3	3	(0.4483)
	Group 4	4	(0.4966)
4. (Complete labels)	Group 1	0	-
	Group 2	1	(0.4819)
	Group 3	3	(0.4135)
	Group 4	6	(0.5803)

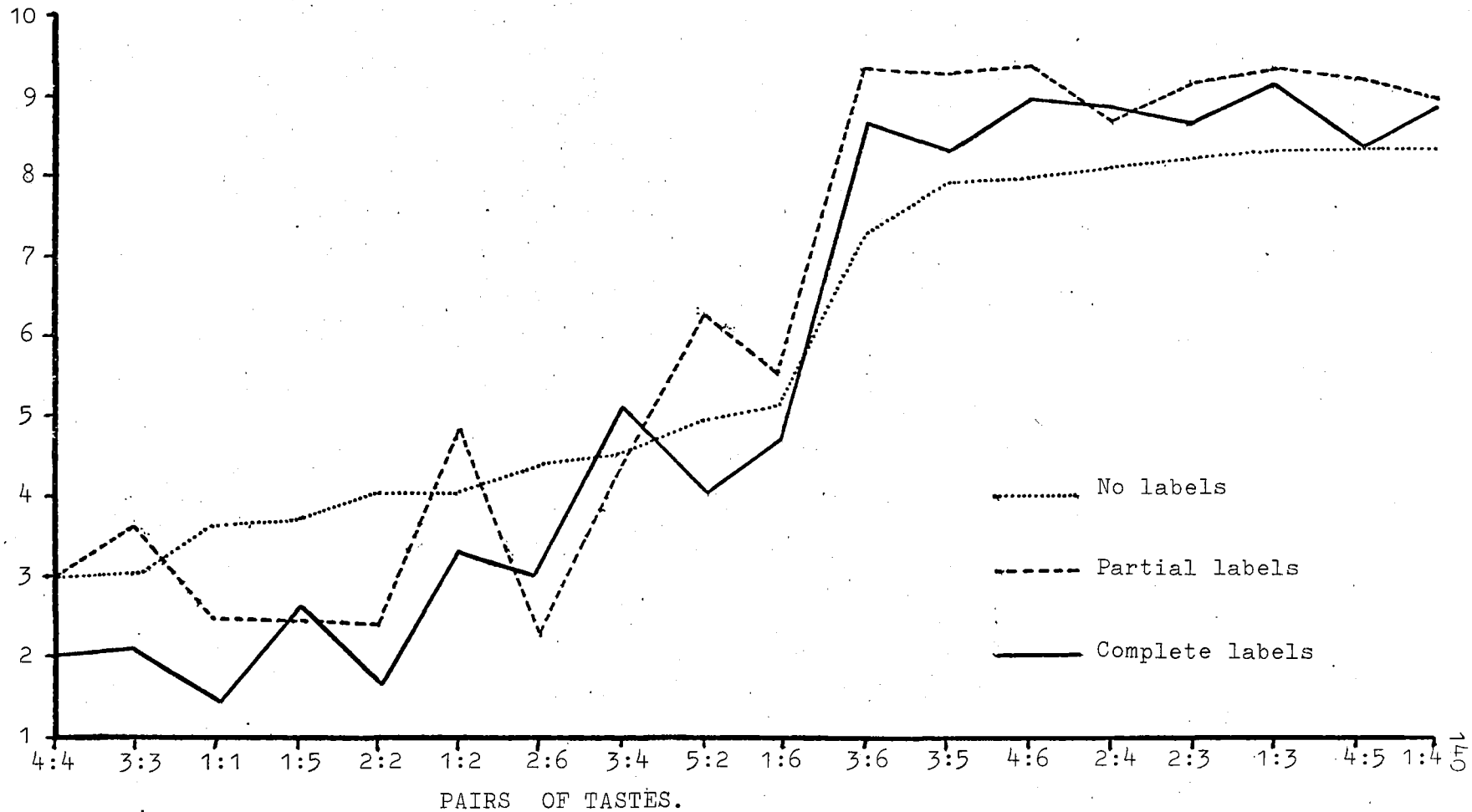
In the majority of cases the a posteriori assignments of subjects to groups corresponds to the actual experimental assignments suggesting that the manipulation of labels has produced fairly consistent influences on the responses generated by individual subjects. Note that the relationship between a priori and a posteriori groupings shows the highest agreement for the salty/sour standard and the poorest correspondence for the D5'G standard. The mean responses made to each pair of stimuli were averaged by pooling across the ten subjects in the four experimental groups and the 20 subjects in the paired comparisons control group. These means are plotted in Figures 4:1 and 4:2. In order to clarify the relationships between the judgements made under each labelling condition, the pairs of tastes have been arranged in decreasing order of similarity (for the responses produced by the "no labels" control group) along the X-axis of the graphs. Figure 4:1 shows the shifts in judgements produced when the "complete labels" and "partial labels" groups are compared with the "no labels" control group. Note that in both of the groups with labels the following general trends emerge:

1. For judgements below a scale value of approximately 5.0 the labelling has enhanced the judged similarity; this enhancement tends to be greater for the "complete labels" group than for the "partial labels" group. Note that in both cases the labels for standards and for comparison stimuli are identical for the same stimuli within each group, and also that of the total set of six stimuli in the "complete labels" group each has a different name



FIG. 4:1

EXPERIMENT 3. EFFECTS OF QUALITATIVE  
LABELS ON TASTE SIMILARITY JUDGEMENTS.

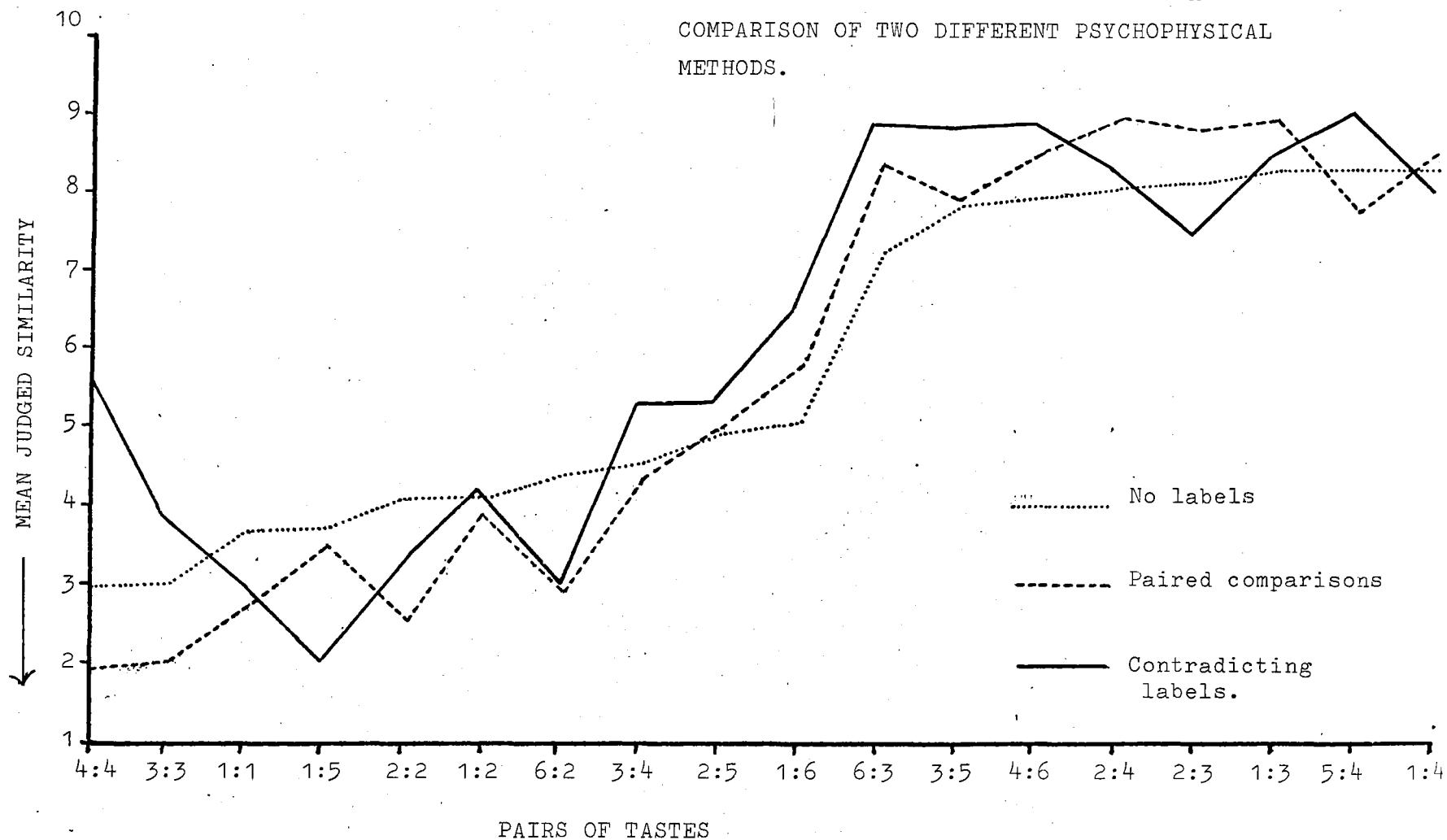


while in the partial labels group two different stimuli are paired with each of the three names. Because of this "doubling up" of names in the partial labels group, we would expect discrimination to be poorer in the partial labels group than in the complete labels group, if labels serve as a cue to sharpen the memory trace and hence facilitate judgement. The graph shows, however, that compared to partial labels, the complete labels have caused similar stimuli to be judged still more similar but they have also caused dissimilar stimuli to be judged more similar, and not more dissimilar as we might expect if the labels were improving the subject's ability to discriminate. The reason for this trend may be revealed by further experimentation, however it should be noted here that the failure of precise labels to enhance the dissimilarity of dissimilar stimuli as compared with the effect of imprecise and confusable labels, does not support the postulate that labels facilitate judgement by improving the discriminability of stimuli. The effect could, however, be related to the choice of stimuli and labels in this particular experimental context. The fact that the "complete label" names mostly list two taste qualities, while the "partial labels" list only a single quality leads to the possibility that stimuli which are completely labelled will have more name-cues in common and the possession of these common cues could enhance similarity between the "completely labelled" stimuli. This point is supported by the fact that most of the individual stimuli at the dissimilarity end of the continuum contain the word sour in one label and bitter in the other in the complete labels

condition, but this is less often the case in the partial labels condition. There is evidence in the pilot study that the sour and bitter qualities are subjectively more similar than sour-sweet, sour-salty, bitter-sweet, and bitter-salty sensations. Confusions between sour and bitter qualities have also been reported by Robinson (1970) who concludes that the difficulty is one of naming - "subjects seem to use the words 'sour' and 'bitter' badly". This fact alone could account for the enhancement of similarity of otherwise dissimilar stimuli which contain the words bitter or sour in both the completely labelled standard and comparison stimulus, but not in one or the other of the partially labelled standard and comparison stimulus.

Figure 4:2 shows the relationship between the paired comparison control group judgements, the "contradicting labels" group, and the "no labels" group. Note that the paired comparisons show an enhancement of similarity and an enhancement of dissimilarity relative to the "no labels" group. Thus paired comparison judgements show the same wider spread of responding along the similarity scale that is produced when stimuli are labelled under the "memorised standards" method. This lends further support to the postulate that trace decay is an important factor in these experiments. When trace decay is minimal, as in paired comparisons, discrimination is sharper than when a temporal delay intervenes between presentation of the unlabelled standards and comparison stimuli. However, the trace decay which occurs as a result of delay between standard and comparison stimulus input can be largely eliminated by adequate

FIG. 4:2 EXPERIMENT 3. EFFECTS OF CONTRADICTING LABELS ON SIMILARITY JUDGEMENTS AND A COMPARISON OF TWO DIFFERENT PSYCHOPHYSICAL METHODS.



pairing of labels with the stimuli which are to be judged against each other. The "contradicting labels" group is also plotted in Figure 4:2. Note that when identical stimuli are given different labels (e.g. 4:4, 3:3 but not, contrary to predictions for 1:1 and 2:2) judged similarity is reduced but when different stimuli are given the same label (e.g. 1:5, 2:6) similarity is enhanced. The data presented in Figures 4:1 and 4:2 on the whole lend a considerable degree of support to the hypothesis that similarity can be manipulated by the appropriate pairing of real word labels with the stimuli.

The mathematical model that we began to develop in Chapter 2 can be applied to much of the data in the present experiment.

Let us redefine the terms as follows: we shall refer to differentiated memory traces as  $D$  and undifferentiated or partially decayed memory traces as  $d$ . The symbol  $s$  refers to judged similarity between standards ( $i$ ) and comparison stimuli ( $k$ ). The baseline used will be the mid-point of our 10-point category scale; that is a scale value of 5.0. The simple model outlined in Chapter 2 states that:

If we accept the finding that nonsense labels reduce memory trace differentiation then the model states that

$$\text{if } D_i \text{ s } D_k > 5.0 \quad \text{then } d_i \text{ s } d_k < D_i \text{ s } D_k$$

$$\text{and if } D_i \text{ s } D_k < 5.0 \quad \text{then } d_i \text{ s } d_k > D_i \text{ s } D_k$$

In Experiment 1 it was noted that 73.1% of the points in Condition 1 and 84.0% of the points in Condition 2 fitted this simple model.

For two tailed probabilities, the binomial test shows that the data for Condition 1 is significant beyond the 0.05 level while the data for Condition 2 is significant beyond the 0.01 level.

The model, in its above form, can be applied to the present data. We have postulated that real-word labels enhance memory trace differentiation. We therefore predict that, compared to the no-labels control group, both the partial labels and the complete labels groups will show increased differentiation. For the partial labels group, 14 out of 18, or 77.7% of the observations fit the model; this is significant at the 0.05 level (2-tailed) on the binomial test. For the complete labels group the proportion is 16 out of 17 or 94.1%; a value which is significant beyond the 0.005 level on the binomial test. A second prediction, that the complete labels group will show greater differentiation than the partial labels group, is not supported by the data, perhaps for the reasons suggested earlier.

We can also postulate that the longer delay in the memorised standards ("no labels") condition as compared with the paired comparisons group will lead to a decay in the memory traces for the former group. On this prediction 16 of the 18, or 88.8% (significant at 0.005 level) fit the model.

The model thus appears to have some generality, though it is weak in the sense that it makes no quantitative predictions of the degree to which differentiation might be expected to occur under various labelling conditions. We will return to this problem in the following chapter.

It was suggested, in the conclusion to Chapter 3, that nonsense labels may, in some way, act as a second stimulus dimension such that subjects judging labelled shapes are responding to the complex stimulus shape-plus-label as a 2-dimensional continuum. The problem of how the labels are combined with the stimuli to produce an overall assessment of similarity needs further clarification.

Two alternatives seem possible: Let us suppose that an effect similar to that proposed by McCormack (1958) to explain the facilitation of motor learning by verbal pretraining operates. That is the labels generate cues ( $l_1, l_2, l_3, l_n$ ) which accompany the stimuli  $S_1, S_2, S_3, \dots, S_n$ . Hence when  $S_1$  is encountered, the total complex which enters into the judgement is  $S_1 + l_1$ . The first alternative is that the subject weights both the stimulus and the label so that when a similarity assessment between  $S_1$  and  $S_2$  is required the subject considers:

- (a) the similarity of  $S_1$  and  $S_2$  and
- (b) the similarity of  $l_1$  and  $l_2$  and

then comes to an overall assessment in terms of the two dimensions involved. The problem for the psychologist then lies in determining how the two dimensions are weighted under various labelling conditions. The second alternative is that the labels are not weighted as a separate dimension, but rather they serve as information carrying cues which aid the subject in his classification of the sensory input which arises from his receptors. In this sense the labels merely increase the probability that

the qualities present in a given stimulus will be detected or attended to by the subject. If we assume that the nonsense labels used in Experiments 1 and 2 are equisimilar pairwise, then Experiments 1 and 2 favour the dimensional interpretation, since meaningless labels were used there but systematic shifts in similarity occurred. The absence of adequate control groups confounds this interpretation however. It may be that the nonsense labels did not decrease the range of responding shown by subjects, but rather the use of self-provided real-word labels in the control conditions increased the range of responding.

To investigate this problem in greater detail it was decided to perform a series of multi-dimensional scalings of the similarity judgements produced by individual subjects in the 4 labelling conditions of this experiment. It was postulated that if, by providing real-word labels for the stimuli, we increase the dimensionality of the stimuli being judged then this increase in dimensionality should appear more frequently as we move from the "no labels" to the "partial", and "complete labels" condition.

The Shepard-Kruskal multidimensional scaling approach was used in the form of TORSCA version 9, a FORTRAN IV program written by Young and Torgerson (1967). The details of the program have been described in Young (1968) and will be very briefly summarised here.

The typical starting point for a multidimensional scaling is a matrix of judged similarities, such as is generated, for example by paired comparisons of the stimuli whose underlying



dimensionality we wish to determine. In general it has been proved by Young and Householder that, given a set of interpoint distances, the dimensionality of that set of points can be determined. Kruskal (1964) has defined multidimensional scaling as the problem of representing  $n$  objects by  $n$  points, so that the interpoint distances are monotonically related to the experimental dissimilarities between objects. Young (1968) points out that the program used here computes a geometric representation of a data matrix such that the distances between the points in the representation best reproduce the order of the entries in the data matrix. This problem was first posed and solved by Shepard (1962). The geometric representation may be in any Minkowski space.

Distances in the Minkowski spaces are calculated according to the formula:

$$d(x,y) = \left[ \sum_{i=1}^m |x_i - y_i|^r \right]^{1/r}$$

where  $x = (x_1 \dots x_m)$  and  $y = (y_1 \dots y_m)$  for  $r \geq 1$ .

When  $r = 2$ , an Euclidian solution results;  $r$  is set at 1 for "city block" space.

However, Hyman and Well (1967, 1968) have suggested that when the stimuli used behave as "unitary wholes" then similarity is best represented by the Euclidian space; it was therefore decided to perform the scalings in Euclidian space (Minkowski constant = 2) in this experiment. The program automatically checks for

and lists violations of triangular inequality and Kruskal's stress (Kruskal, 1964) is calculated as a measure of the goodness of fit of the derived configuration to the original similarity measures.

The data were read into the computer as a series of triangular half-matrices, one for each subject. In this experiment we have 4 standard tastes judged against the same 4 tastes as comparison stimuli (plus an additional two comparison stimuli which are not included in the analysis). The self comparisons, or judgements between a given taste as standard and the same taste as comparison stimuli do not usually form part of the data analysed by multidimensional scaling. However, there are two important considerations regarding the present experiment. First, the self-comparisons, by inspection, appear to be the judgements that are most heavily affected by the labelling manipulation; secondly although the 4 standards and 4 of the comparison stimuli are physically identical in taste, they are in some cases psychologically different stimuli because there is not always a one-to-one pairing between tastes and labels. Therefore to maximise the likelihood of showing a shift in the underlying dimensionality of the stimuli when different sets of labels are paired with the tastes, it is necessary to treat the 4 standards and the 4 corresponding comparison stimuli as different stimuli, thus producing an 8 x 8 matrix of responses. (Stimuli 5 and 6 do not enter into the analysis since they were not used as standards and were not compared in turn with every other

stimulus in the experiment.)

The original 4 x 4 whole matrix was broken down into an 8 x 8 half matrix as follows:

Original Matrix

	Standard 1	Standard 2	Standard 3	Standard 4
c.s.1	a	b	c	d
c.s.2	e	f	g	h
c.s.3	i	j	k	l
c.s.4	m	n	o	p

New 8 x 8 Matrix

	A	B	C	D	E	F	G	H
	Std.1	c.s.1	Std.2	c.s.2	Std.3	c.s.3	Std.4	c.s.4
Std.1 A								
c.s.1 B	a							
Std.2 C	-	b						
c.s.2 D	e	-	f					
Std.3 E	-	c	-	g				
c.s.3 F	i	-	j	-	k			
Std.4 G	-	d	-	h	-	l		
c.s.4 H	m	-	n	-	o	-	p	

There are 12 empty cells or "passive elements" in this half-matrix. These unfilled cells are replaced in each case by the mean of the 16 "active elements" in the filled cells before the multidimensional scaling begins.

Since there are a relatively large number of passive cells in the original dissimilarities (or data input) matrices, the stress calculated for each solution is ultra-conservative and

hence misleading. A measure which corresponds approximately to stress can be obtained by calculating the percentage of violations of triangular equality which occur in the disparities matrix corresponding with each n-dimensional solution. The rejection of any solution for which the violations exceed  $7\frac{1}{2}\%$  is suggested as a reasonably satisfactory cutoff point. A scattergram was drawn to show the dependence of stress on percent violations in the disparities matrix as calculated in 1, 2 and 3 dimensional solutions from 10 subjects responding under conditions not unlike these. There were no passive cells in the matrices plotted. The graph suggested that at least when the violations did not exceed 10% the stress was generally not greater than 0.10; i.e. a "fair" goodness of fit according to Kruskal's (1964) evaluation. There is evidence, however, that when the number of stimuli is small, Kruskal's evaluation is not sufficiently stringent (Klahr, 1969).

Nine of the ten subjects in the "contradicting labels" condition were analysed in 1, 2, 3, 4 and 5 dimensional space; however it was evident that two, or perhaps 3 dimensions were adequate to reasonably conceptualise the subject's responses. The remaining subjects were therefore analysed in 1, 2 and 3 dimensions, only. Of the 40 3-dimensional solutions only 3 were associated with violations greater than 0.0%. This indicates a

very low stress for the majority of the solutions which in turn suggests that 2 or 1 dimensions would provide adequate solutions. The two dimensional solutions provided a more satisfactory set of results; the 1 dimensional solutions were consistently associated with high percent violation figures. (Only 4 of the 40 solutions were associated with violations of less than 25%.) It was therefore decided to accept the 2-dimensional solutions as being likely to provide the most parsimonious conceptualisation of the underlying dimensionality.

A priori, if labels alter the underlying dimensionality of the stimuli, then we would expect the most marked changes in dimensionality to appear in those cases where the qualities listed in the descriptive labels were changed by the experimenter from one treatment condition to the next. By assuming that symmetry holds, it was predicted that the following pairs of stimuli would appear as one of the poles to a dimension equally often in the "no-labels" group, since they represent symmetrically opposite judgements in the original 4 x 4 matrix:

Stimuli 1 and 2, AD and BC

Stimuli 1 and 3, AF and BE

Stimuli 1 and 4, AH and BG

Stimuli 2 and 3, CF and DE

Stimuli 2 and 4, CH and DG

Stimuli 3 and 4, EH and FG

It was assumed that any large deviations from the frequencies of occurrence of the above symmetrical pairs as poles to a

dimension which occurred in the partial, complete and contradicting labels groups could be attributed to the labels that were paired with each stimulus.

However, the tables below show that symmetrically opposite pairs of stimuli did not tend to emerge as dimensional poles equally often in the "no-labels" group. In fact the high level of intersubject variability which emerges makes it difficult to draw any conclusions about the specific changes in the underlying dimensionality of the stimuli which can be attributed to labelling.

Tables 4:3 and 4:4 list the frequencies with which each pair of stimuli emerged as opposite poles to a dimension in the two dimensional solutions only. In Table 4:3 the frequencies with which each pair of stimuli emerged as the two most widely separated poles on any dimension are listed. A less stringent criterion was taken in Table 4:4. In this case a tally was made of all pairs of stimuli which emerged on any dimension with a varimax rotated configuration value which was greater than  $\pm 0.5$ . The varimax criterion for rotation of axes has been described in detail by Kaiser (1958).

In the occasional rare cases where two stimuli emerged with equal separation at a given pole of a dimension, both of the stimuli are listed.



Notice that the labelling appears to have affected the underlying dimensionality of the stimuli most markedly in the contradicting labels (K.L.) group. The overall pattern, however, shows a strong consistency across groups, suggesting that any effect of labels on the underlying dimensionality of the stimuli is very definitely second order.





The multidimensional scaling was carried out in the attempt to show that labelling might induce changes in the underlying dimensionality of the responses made to the 4 tastes. The data however provide little evidence for such a change, in fact there is a high degree of uniformity in the dimensions that were extracted from subjects in the four labelling groups. One further peculiarity noted in the two dimensional solutions was that approximately half of the solutions in each labelling condition showed a marked tendency for four distinct clusters of points to emerge - (AC), (FH) on one dimension and (BD), (EG) on the other. However, since the stimulus pairs AB, CD, EF and GH are in fact physically identical, and in most cases (except the contradicting labels condition) were labelled identically it would appear that the extraction of a second dimension is actually artefactual and that, although the percentage of triangular equality violations were high in the 1-dimensional solutions, these latter analyses more correctly reflect the underlying perceived dimensionality of the stimuli. Klemmer and Shrimpton (1963) using the Shepard (1962) scaling technique found that a configuration containing two clusters, with large between-cluster and small within-cluster distances will collapse into one dimension with zero within-cluster distances and large between-cluster distance. A similar effect has been reported by Rankin, Markley and Evans (1970) and it appears likely that the present data could likewise be collapsed into a smaller dimensionality by combining clusters of points, although the clusters reported here are not directly comparable with those

mentioned above. Both the 1-dimensional and the 2-dimensional solutions for the majority of subjects suggest that stimuli ABC and D are one pole and stimuli EFG and H are the opposite pole of a unidimensional continuum of salty/sourness - bitter/sweetness or perhaps strong (concentrated)-weak (diluted) tastes, or even unpleasant-pleasant tastes. In fact, if we take the most widely separated poles on the 1-dimensional solutions, and neglect the high percentage of violations for these solutions, 10 of the 10 solutions in the "no labels" and "partial labels" conditions and 9 of the 10 solutions in the "complete labels" condition favour a division of the stimuli into the two poles ABCD and EFGH. This dichotomous division of the stimuli appears to have been so marked that the labels produced no shift in the underlying dimensionality. One further point remains to be investigated however. Perhaps the labels show a consistent trend towards increasing the distances or degree of separation of the two poles on this single dimension. That is, as the precision of the labels increases, the range of responses should increase and this may be reflected in the degree of separation between labelled stimuli that is shown in the analysis.

For the 1-dimensional solutions, the following pairs of points exceeded  $\pm 0.5$  in the derived configurations:

NO LABELS CONDITION

Subject 1 BF BH DF DH  
 Subject 2 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 Subject 3 AF AH CF CH  
 Subject 4 -  
 Subject 5 AE AG CE CG  
 Subject 6 AE AF AG BE BF BG CE CF CG DE DF DG  
 Subject 7 AF AH BF BH CF CH  
 Subject 8 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 Subject 9 AG AH  
 Subject 10 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 (Total 80 pairs, all favouring the ABCD-EFGH dichotomy)

PARTIAL LABELS CONDITION

Subject 1 -  
 Subject 2 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 Subject 3 AE AF AG BE BF BG CE CF CG DE DF DG  
 Subject 4 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 Subject 5 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 Subject 6 AE AF AG AH BE BF BG BH CE CF CG CH CE DF DG DH  
 Subject 7 AE AF AG AH BE BF BG BH  
 Subject 8 BE BF BG CE CF CG DE DF DG  
 Subject 9 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH  
 Subject 10 CG CH  
 (Total 111 pairs, all favouring the ABCD-EFGH dichotomy)

COMPLETE LABELS CONDITION

Subject 1 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 2 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 3 CE CF

Subject 4 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 5 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 6 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 7 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 8 AE AF AG AH BE BF BG BH DE DF DG DH

Subject 9 AF AG AH BF BG BH CF CG CH DF DG DH

Subject 10 AB AF AH DB DF DH

(Total 128 pairs of which 126 favour the ABCD-EFGH dichotomy)

CONTRADICTING LABELS CONDITION

Subject 1 AD AE BD BE

Subject 2 BE BF BH CE CF CH DE DF DH

Subject 3 AE AF AG BE BF BG CE CF CG

Subject 4 CE CF CG DE DF DG

Subject 5 AE AF AG AH BE BF BG BH CE CF CG CH DE DF DG DH

Subject 6 BE BF DE DG

Subject 7 BF BH DF DH

Subject 8 AH BH

Subject 9 AF EH GH

Subject 10 BG DG

(Total 59 pairs of which 57 favour the ABCD-EFGH dichotomy)

There is thus a clear trend for a stronger separation of stimuli to occur along the single dimension as the precision of the labels increases. The contradicting labels have led to a poorer separation, but the underlying dimensionality remains much the same. We can tentatively conclude that labelling does not alter the dimensionality of the judgement task when real word labels are used, but rather the labels alter the perceived relationship of stimuli to each other within the dimension.

The conclusion that a unidimensional classification of the taste stimuli appears to have operated here suggests that the labels are not responded to as a new dimension but rather they simply enhance the distinctiveness of the stimuli within the dimension. This does not support the conclusion drawn in Experiment 2 where it was proposed that nonsense labels increase the dimensionality of the subject's task by providing an additional dimension of uniformly similar verbal cues.

One alternative to the hypothesis that labels modify similarity judgements by adding an extra dimension to the task, is the "overwriting" hypothesis which was developed in Chapter 2. Although Experiments 1 and 3 are not directly comparable, it does appear that the use of real word labels has produced a larger shift in judgements than the use of nonsense labels. This suggests that the meaningfulness of the label is a relevant factor, and that overwriting, if it does occur, is more effective when we overwrite using a real word than when we use a nonsense label. It was suggested that ease of overwriting may be related to the

familiarity of the stimulus where familiarity is defined in terms of degree of prior exposure. However, with taste stimuli we are generally only able to infer that one taste sensation will have been experienced more frequently than another.

An alternative to the overwriting hypothesis is to explain the facilitation of judgement in terms of the provision of additional relevant information by the labels that are paired with stimuli. Presumably if labels merely provide additional and in some cases redundant information then familiarity of the stimuli will be a less critical consideration.

In summary, the greater differentiation or precision of judgement that was observed under the labelled conditions supports the hypothesis that labels facilitate judgement by supplying the subject with additional cues which label aspects of the stimuli that might otherwise have passed unnoticed by the subject, or which may have been noticed but lost due to rapid decay in the memory traces which are generated when a physical stimulus is experienced.

The effect of labelling on the actual judgement process also needs to be considered in some detail. Newell (1968) has described judgement (the act of judging) as a cognitive process with the following characteristics:

1. The main inputs to the process, that which is to be judged, are given and available; obtaining, discovering, or formulating them is not part of judgement.

2. The domain of the output - the set of admissible responses - is simple and well defined prior to the judgement. The response itself is variously called a selection, estimation, assertion, evaluation or classification (in the sense of identification of class membership, not of creating the classes), depending on the nature of the domain.
3. The process is not a simple transduction of information; judgement adds information to the output.
4. The process is not simply a calculation, or the application of a given rule.
5. The process concludes, or occurs at the conclusion of, a more extended process (the causal role is not completely clear).
6. The process is rather immediate, not being extended in time with phases, stages, subprocesses, etc. (If such occur, they tend to be referred to as preparation for judgement.)
7. The process is to be distinguished from searching, discovering, or creating, on the one hand; and from musing, browsing, or idly observing on the other.

In the unlabelled condition of Experiment 3, the judgement task can be considered to possess all seven of the characteristics listed above. However, in the "complete labels" condition, and particularly in those cases where a given stimulus<sup>is</sup>/judged against itself, the task appears to become more of the type described by Shelly and Bryan (1964, p.9) as "the reading of a number off a dial"; a reaction to which the term "judgement" is not considered



to apply. Thus by labelling the stimuli it could be argued that we have changed the judgement task in almost all of its characteristics:

1. The main inputs to the process are altered, however this is more relevant to the question of discrimination than to the problem of judgement.
2. The domain of the output changes from a response based mainly on selection or estimation, to a classification response.
3. The process may move nearer to being a simple transduction of information.
4. Likewise the response could now be considered to involve the application of a rule, particular in the case of judgements between identical stimuli.
5. The process may be less extended than before, since labels appear to facilitate judgement.
6. This thesis is partly concerned with whether or not a verbal labelling or categorisation subprocess occurs during the preparation for judgement. If such a process does occur, then we will have altered not only the characteristics of the judgement task but also at least one of the stages or processes involved in the preparation for judgement.
7. The process could be considered to have moved from the searching end of the continuum to a process more like observation, or choice.

This argument assumes, however, that:

- (a) the subject actually accepts the labels at face value and does not question their accuracy, and
- (b) the judgement is based at least in part on the information conveyed by the labels.

Bock and Jones (1968) p.4. take the view that judgement is a comparison made by a subject between two stimuli each of which are experienced as a continuously variable physical event along a sensory continuum. They add that "When the subject is required to choose among alternatives on the basis of personal preference, his response is ordinarily called a choice."

If we consider Newell's description of the judgement process the task remains judgemental only in the sense that the subject is still required to order the labelled stimuli along a dimension of similarity. However, on the Bock and Jones definition, the task retains the characteristics of a judgemental rather than a choice process.

## CHAPTER V

### SOME EFFECTS OF VERBAL LABELS ON NONSENSE SHAPES.

In the experiments described so far, the labels which were overtly paired with stimuli by the experimenter either provided no additional information about the physical characteristics of the stimuli (e.g. the nonsense labels of experiments 1 and 2) or else they provided relevant information in varying degrees of redundancy, or even contradicting or incorrect cues (Experiment 3). The real word labels of experiment 3 were not designed to attempt to alter or modify the input associated with the taste stimuli, rather they were provided as supplementary or additional input cues for the subject to use as he chose. It was noted that the familiarity (as shown by degree of prior exposure) of the stimuli was a relevant, but uncontrolled variable.

Since similarity judgement can sometimes tend to nothing more than a categorising process, it would appear to be likely that the meaningfulness or ease with which a stimulus can be categorised is a relevant factor in similarity judgement. Meaningfulness has frequently been associated with the number of associations that are aroused by a stimulus (e.g. Osgood, Suci and Tannenbaum, 1957) or have been conditioned to a stimulus (e.g. Staats, 1968). A stimulus which arouses few associations is thus expected to be relatively meaningless compared with a stimulus which arouses many

associations within a subject.

It has been suggested that labels enhance the judged similarity and differences between stimuli by several possible mechanisms including:

- (a) the reduction of trace decay when there is a delay between presentation of the standard and comparison stimuli,
- (b) the channelling of the subject's perceptual response towards certain relevant features of the labelled stimulus and hence decreasing generalization between stimuli, and
- (c) the labels may increase the dimensionality of the stimulus input so that instead of judging similarity between stimulus 1 and stimulus 2, the judgement is made between stimulus 1 + label 1 and stimulus 2 + label 2.

If the labelling shift is entirely attributable to trace decay then by presenting pairs of labelled stimuli to the subject we would not expect to observe the labelling shift for stimuli that are paired with real-word labels which carry redundant sensory information. If labels have an effect other than reducing trace decay then with simultaneous presentation of stimuli a shift in judgement should still be observed. Furthermore, in the experiments previously described, subjects were never instructed to refer to the labels as part of the overall stimulus complex which was to be considered during judgement. By actually instructing the subjects to assign some weight to the labels as well as the stimuli it may be possible to force subjects to use the labels as an additional dimension. By having a further group





of subjects judge the similarity in meaning of the labels used it may be possible to determine, through multidimensional scaling whether the labelled stimuli are responded to mainly in terms of their physical characteristics (as shown by a response pattern which is similar to that generated by subjects who are required to judge physical (unlabelled) similarity only) or in terms of their labelled characteristics (i.e. a response pattern which is like that produced by subjects judging the meaning of the labels only). An answer to this problem would give us information regarding the question of whether stimuli are cognitively manipulated in terms of raw sensory input or in terms of some kind of simplified verbal schemata in which stimuli are classified by the names which subjects assign to them. The stimuli chosen for this experiment were selected from the Vanderplas and Garvin (1959) list of nonsense shapes. The nonsense shapes have been classified in terms of complexity by the number of randomly plotted points that were joined to make each solid figure. This experiment uses the 4 highest and the 4 lowest association value shapes from the 24 point series of shapes, that is, stimuli 1 to 4 (high A.V.) and stimuli 26 to 30 (low A.V.) as shown in Vanderplas and Garvin (1959). The stimuli are shown in actual size in Figure 5:1.

It was postulated that if labelling aids the assessment of similarity and the classification of stimuli by sharpening discrimination (or reducing generalization) then easily labelled (or high association value) shapes should be judged as less similar to each other than low association value shapes which are





FIG. 5:1

The 8 Vanderplas and Garvin (1959) Nonsense  
Shapes Used in Experiments 4 and 5.

High Association Value Shapes

Shape				
Stimulus Number	1(A)	2(B)	3(C)	4(D)

Low Association Value Shapes

Shape				
Stimulus Number	5(E)	6(F)	7(G)	8(H)

a priori difficult to label subjectively. Since the stimuli are to be presented simultaneously, it will not be possible to attribute any shift in judgement to trace decay.

Because associations can be expected to vary considerably from one subject to the next, it was decided to assess independently the association value of each shape in the experiment for each subject. We would expect that those subjects who are required to assess the 'meaningfulness' of a set of shapes before judging the similarity of the same set of shapes to each other will be more aware of the differences between high association value shapes and the similarities between low A.V. shapes than subjects who judge similarity between the shapes first and then assess the 'meaningfulness' of the stimuli. We would expect to find little or no difference between the judgements made by the two groups on the low association value shapes if awareness of 'meaningfulness' of the shapes is a relevant variable in this experiment.

#### EXPERIMENT 4.

##### Stimuli and Procedure.

The stimuli were the Vanderplas and Garvin (1959) 24 point shapes, numbers 1 to 4 and 27 to 30. A photocopy of each shape was mounted on a white card measuring 2 by 4 inches. To introduce the subject to each stimulus the entire set of 8 shapes were first exposed to the subject for 10 seconds each. A set of test cards was then shown in which one of the target stimuli was placed randomly in an array of 9 other masking stimuli selected from the 16 and 24

point shapes. The association test consisted of a form on which subjects were given as much time as they required to write a word or phrase describing whatever they were reminded of by the 8 shapes.

The instructions were as follows:

"INSTRUCTIONS FOR ASSOCIATION TEST.

I am going to show you a number of shapes. Some of the shapes may remind you of some familiar object or situation while others may not remind you of anything. Your job will be to name whatever the shape reminds you of, if anything. If the shape reminds you of something that you can describe in a word or two, simply write that word or phrase in the appropriate space below. If the shape doesn't remind you of anything, write "nothing". It is essential that you write something - either a word, or phrase if the shape reminds you of something you can describe, or else "nothing" if you are not reminded of anything during the 10 seconds for which I will show you each shape."

The subjects were all undergraduate volunteer students and no subject served more than once under any of the 4 conditions. There were 10 subjects in each of the 4 experimental conditions.

Condition 1.

The subjects were first shown the 8 shapes, exposed consecutively for 10 seconds each. The identification test cards were then shown, subjects having been previously warned that they would be required to pick the 8 target shapes out of an array of masking shapes. The subject was told that only one of the shapes on each card had been seen before and was required to guess if uncertain (unlimited time



was given). The responses were scored by the experimenter as "correct" or "incorrect".

The 8 stimuli were then shown again, while subjects wrote their association for each shape on the form provided. Fifty-six paired comparison similarity judgements were then taken, judgements were made using the same 10 point scale (with numerals) as in Experiment Two. The pairs of stimuli were presented in random order, with all stimuli serving once as standard and once as comparison stimulus against which the 7 remaining stimuli were judged. Subjects wrote their chosen response in the appropriate space below the trial number on their instructions form. The instructions to subjects for the judgement tasks in Conditions 1, 2, 3 and 4 were as follows:

"INSTRUCTIONS FOR CONDITIONS 1 AND 2 (SHAPE SIMILARITIES)

I am going to show you a number of shapes - your job is to judge the similarity of certain pairs of shapes, using the scale

1. Identical
2. Nearly identical
3. Very similar
4. Similar
5. Slightly similar
6. Slightly dissimilar
7. Dissimilar
8. Very dissimilar
9. Extremely dissimilar
10. Completely different

Here is a sample of the kinds of shapes you will be comparing: The shapes will be presented to you in pairs; you are required to judge how similar you think the second shape in each pair is to the first member of the pair. For example, if you think the second shape in the pair is IDENTICAL to the first shape, write the number 1 in the appropriate space below; if you think it is COMPLETELY DIFFERENT write 10, etc. It is essential that you avoid repeating the same response for each different pair, however, try to respond as quickly and as accurately as possible."

"INSTRUCTIONS FOR CONDITION 3 (LABELLED SHAPE SIMILARITIES).

I am going to show you a number of shapes - your job is to judge the similarity of certain pairs of these labelled shapes, using the scale:

1. Identical
2. Nearly identical
3. Very similar
4. Similar
5. Slightly similar
6. Slightly dissimilar
7. Dissimilar
8. Very dissimilar
9. Extremely dissimilar
10. Completely different.

The shapes, along with their appropriate labels, will be presented to you in pairs; you are required to judge how similar you think the second labelled shape in each pair is to the first member of the

pair. For example, if you think the second labelled shape in the pair is IDENTICAL to the first shape, write the number 1 in the appropriate space below; if you think it is COMPLETELY DIFFERENT, write 10, etc. Try to make your judgements by combining the information given in both the physical shape and the verbal label, to give an overall judgement of similarity. That is, don't judge either the shapes only, or the labels only, against each other, but try to combine both shape and label into a single overall judgement for each pair of cards.

Try to respond as quickly and as accurately as possible."

"INSTRUCTIONS FOR CONDITION 4 (LABEL SIMILARITIES).

I am going to show you a number of words - your job is to judge the similarity in meaning of certain pairs of words, using the scale

1. Identical
2. Nearly identical
3. Very similar
4. Similar
5. Slightly similar
6. Slightly dissimilar
7. Dissimilar
8. Very dissimilar
9. Extremely dissimilar
10. Completely different.

The words will be presented to you in pairs; you are required to judge how similar you think the meaning of the second word in each

pair is to the first member of the pair. For example, if you think the second word in the pair is IDENTICAL in meaning to the first word, write the number 1 in the appropriate space below; if you think it is COMPLETELY DIFFERENT in meaning, write 10, etc. It is essential that you avoid repeating the same response for each different pair, however, try to respond as quickly and as accurately as possible."

#### Condition 2.

Condition 2 was identical to condition 1 except that the association test followed the similarity judgement task rather than preceding it. Conditions 1 and 2 form a complete experiment in that any systematic differences in judgement across the two groups of subjects can be attributed to the fact that subjects in condition 1 had overtly assigned names or labels to the nonsense shapes, while subjects in condition 2 had not named the stimuli before making their judgements.

#### Condition 3.

For condition 3, the sequence of events was the same as in condition 1. However, the associations made by subjects in conditions 1 and 2 were examined, and a single word which reflected the most frequently occurring association to each of shapes 1 to 4 was printed clearly in black ink above the appropriate shape on the cards used in the similarity judging task. Instructions to subjects were modified to emphasise that subjects should attempt to consider both the label and the physical shape and combine both elements into a single overall similarity judgement. The labels

used for stimuli 1 to 4 were TREE, BEAR, SWAN and CRAB while stimuli 5 to 8 (i.e. the Vanderplas and Garvin shapes 27 to 30) were all labelled with the word NOTHING.

Condition 4.

The 10 subjects in this condition were simply instructed to judge the similarity of the meanings of the 5 words TREE, BEAR, SWAN, CRAB and NOTHING. The words were printed as before on white cards measuring 2 by 4 inches and the total set of 20 paired comparisons were presented in random order. The nonsense shapes were never viewed by these subjects.

RESULTS.

Conditions 1 and 2 test the hypothesis that stimuli which are relatively meaningful in the sense that they can be readily labelled by a subject will be judged as less similar to each other than relatively meaningless shapes which are not easily labelled by a subject. They also test the hypothesis that a prior task in which subjects are required to label the stimuli will make subjects more aware of the stimulus differences so that discrimination or judged dissimilarity between pairs of stimuli will be greater for those subjects who label the stimuli first (in the association test) than for those subjects who judge similarity first, then label the stimuli.

Conditions 1 and 2 also provide a check on the possibility that as the experiment progresses, all stimuli may have come to acquire meaning, due to longer exposure to the shapes. In fact,

the 10 subjects in condition 1 gave a total of 25 responses of "nothing" while in condition 2 the total number of occasions on which a stimulus shape failed to arouse an association was 30, suggesting that the longer exposure to the shapes that was received by subjects in condition 2 (during the judgement task) did not increase their probability of producing an association to each shape in the association test.

The data showed no systematic trend for condition 1 subjects to show a greater range of scale usage than the subjects in condition 2. The hypothesis that subjects who judge similarity first and then label the shapes will be less aware of shape similarities and differences than subjects who label the stimuli first and then judge similarity was, therefore, not supported.

Our other hypothesis is that the pairs of stimuli which have high association values will be judged as more dissimilar to each other than pairs of stimuli which have low association values. We can test this hypothesis by partitioning subjects according to their actual association-test responses. Thus, for each stimulus pair we can pool data from subjects into one of 3 categories:

- (a) both stimuli in the pair produced associations; these will be referred to as the high/high group,
- (b) neither of the stimuli in the pair produced as association (the low/low group), and
- (c) one of the stimuli produced an association, the other member of the pair did not (referred to as the high/low group).

Unfortunately, the majority of subjects gave associations to

more than half of the shapes; we thus have an abundance of high/high paired comparisons but a shortage of low/low comparisons. In fact, comparisons between the mean judgements produced in all 3 of the response classes are possible only for the stimulus pairs 4:5, 4:6, 4:7, 4:8, 5:6, 5:7, 5:8, 6:7, 6:8, and 7:8. The pooled data from groups 1 and 2 (i.e. association tests both preceding and following the judgement task) are shown below:

Stimulus Pair	High/High	High/Low	Low/Low
4:5	4.41	5.70	3.50
4:6	5.06	5.12	5.25
4:7	5.30	6.80	6.00
4:8	7.18	7.77	7.50
5:6	4.66	6.83	4.71
5:7	5.58	7.15	5.16
5:8	5.37	7.50	6.32
6:7	7.00	6.96	6.29
6:8	7.00	7.50	4.45
7:8	4.43	5.20	5.16

In only 4 of the 10 stimulus pairs the high/high judgement exceeds the low/low value; our hypothesis is thus not supported. In 18 cases out of 20, however, the mean judgement for the high/low group exceeds either the high/high or the low/low mean judgement. There is thus a tendency for both the high/high and the low/low judgements to show greater similarity for a given pair of shapes than when one shape was of high and the other was of low association value. One possible explanation for this is suggested by the actual

associations given to the stimuli. An independent scorer tallied the entire set of 30 association test responses (from conditions 1, 2 and 3) in terms of animate, inanimate and "nothing" responses. The percentages of responses which fell into each category for each stimulus were as follows:

Stimulus	Animate	Inanimate	Nothing
1 (A)	60.00%	36.67%	3.33%
2 (B)	86.67%	0.00%	13.33%
3 (C)	96.67%	3.33%	0.00%
4 (D)	66.67%	6.67%	26.67%
5 (E)	36.67%	6.67%	56.67%
6 (F)	36.67%	16.67%	46.67%
7 (G)	3.33%	23.33%	73.33%
8 (H)	33.33%	26.67%	40.00%

Thus for the high association value shapes (stimuli 1 to 4) the majority of responses were animate (in fact, vertebrate animals, particularly mammals) while for stimuli 5 to 8 the majority of responses were "nothing" (meaning that the shape did not remind the subject of "any familiar object or situation"). We would thus expect that, if labelling influences similarity judgement, the responses made between high association value stimuli would in fact be influenced by labels referring to a highly homogeneous set of objects (vertebrates) while for the low association value shapes the judgements are again made between stimuli which are associated with a homogeneous set of labels - in this case, the verbal response "nothing". For judgements between the high/low stimulus pairs,



however, the subject would, in general be comparing the category "nothing" with the category "vertebrate animal" hence we would expect an enhancement in the judged dissimilarity between the stimuli under this situation. The 2-tailed probability of obtaining 18 out of 20 differences in the same direction by chance is given by the Binomial test as less than 0.01, the observed trend in the data is thus statistically significant. The hypothesis that subjects who see no associations in a pair of nonsense shapes will judge them to be more similar than subjects who do see familiar objects or situations in the same pair of shapes has not been supported. However, an alternative hypothesis which states that there will be greater judged dissimilarity between meaningful and meaningless shapes than between either pairs of meaningful, or pairs of meaningless shapes was supported by the data. This latter finding lends further support to the theory that labelling influences judgement since, by manipulating the association value of nonsense shapes we have been able to produce shifts in judgements across groups of subjects which are, presumably, independent of the actual physical contours of the stimuli used. It appears then that meaningfulness of stimuli (as measured by the number of associations aroused) is a relevant factor in the judgement task; the meaningfulness factor appears to exert its influence on subject's judgements even in the absence of instructions which suggest to the subject that he judge similarity in terms of both the physical shape and the associations evoked by the shape. That is, subjects appear spontaneously to make use of the cues that arise from the

associations evoked by the shapes and these associations enter into the judgement process as additional stimulus cues which are considered to be relevant to the assessment of similarity.

A second reason for performing the present experiment was to attempt to determine whether the labels that are paired with stimuli act as an additional or secondary cue or whether they are assimilated with the physical stimuli along already existing dimensions. That is, when we pair labels with stimuli we may either

- (a) add a new dimension to the input, as was suggested as an explanation for the shifts in judgements that were apparently induced by nonsense labels in experiments 1 and 2, or
- (b) we may simply increase the magnitude of the input along already existing dimensions as appears to have happened in experiment 3 where the labels salty, sour, etc., seem to have simply increased the subject's awareness or certainty that these qualities were actually present in the stimuli to which he responded.

It may be that both factors operate under different circumstances, depending on what form the labels take, and on what new or additional information they give to the subject.

One major characteristic of the present set of nonsense shape stimuli is that they are subjectively difficult to order along a continuum of similarity on the basis of their physical characteristics alone. (Compare the nonsense stimuli with the shapes used in experiment 2 where the underlying dimensionality was immediately obvious.) Consider now the problems faced by a subject who is required to classify these heterogeneous and complex shapes in

terms of their similarity. In the absence of any marked physical basis for classifying the stimuli it seems reasonable to postulate that he may classify them in terms of real-world physical objects which the shapes remind him of. If this were so, then we would expect the dimensionality of the responses made by subjects to the shapes to be reflected to some extent in the dimensionality of the responses made by subjects to the labels that are paired with the shapes. If, however, shape similarity judgements are made on purely physical terms and independently of the similarity of the associations aroused by the shapes then we would expect the dimensionality of the labels to differ from the dimensionality of the shapes and the dimensionality of the responses made to the overtly labelled shapes in condition 3 should reflect both the dimensionality of the shapes and of the labels. In other words, if labels paired with stimuli add a new (verbal) dimension to the judgement task, then the dimensionality of the shape plus label judgements should in some way emerge as the sum of the dimensionalities of the judgements made to the shapes alone and to the labels alone. If, however, the labels simply assimilate with already existing dimensions (as we might expect if covert labelling of shapes occurs regardless of our overt label manipulation) then the dimensionality of the shape judgements should not differ greatly from the dimensionality of the shape-plus-label judgements. (We will tentatively assume that the overt labels which we paired with the stimuli do not differ greatly from the covert labels which subjects would have associated with the stimuli in the absence of

our verbal manipulation.)

The similarity matrices generated by each of the 40 subjects in conditions 1 to 4 were analysed using TORSCA 9, the FORTRAN IV program by Young and Torgerson (1967) which was described in chapter 4. Since input was in the form of triangular half-matrices with no passive elements, the stresses calculated by the program for each solution provide a valid index of goodness of fit. Data from each subject in conditions 1 to 3 was analysed in 1, 2 and 3 dimensions for an 8 x 8 half-matrix of judged dissimilarities. The 5 x 5 half-matrices for condition 4 were analysed in 1 and 2 dimensions only. Hoben (1968) performed a nonmetric scaling of 20 random shapes (which were similar to those used in this experiment) in various Minkowski spaces. In contrast to the approach used here, he analysed pooled data from groups of more than 200 subjects. There was a clear trend for the data to fit an Euclidean model as well as, or better than, various other Minkowski metrics. The poorest fit was obtained with the city-block metric. It was therefore decided to perform the scalings for this experiment in Euclidean space.

Stresses tended to be rather high for all solutions, perhaps due to the higher level of intrasubject response variability that was noted in this experiment as compared with the earlier studies. Table 5:2 shows the stresses associated with each solution for all 4 experimental conditions:

TABLE 5:2

Subject		1	2	3	4	5	6	7	8	9	10
Condition 1	3D Soln.	0.056	0.012	0.047	0.022	0.031	0.064	0.056	0.047	0.069	0.058
	2D Soln.	0.167	0.170	0.085	0.012	0.047	0.158	0.137	0.064	0.133	0.133
	1D Soln.	0.393	0.404	0.370	0.161	0.166	0.278	0.248	0.239	0.267	0.336
Condition 2	3D Soln.	0.039	0.057	0.068	0.053	0.126	0.066	0.043	0.079	0.062	0.028
	2D Soln.	0.123	0.128	0.145	0.156	0.218	0.138	0.151	0.152	0.173	0.092
	1D Soln.	0.301	0.231	0.315	0.301	0.376	0.269	0.275	0.275	0.366	0.247
Condition 3	3D Soln.	0.064	0.026	0.094	0.086	0.066	0.029	0.030	0.059	0.027	0.089
	2D Soln.	0.124	0.117	0.177	0.165	0.152	0.108	0.084	0.147	0.093	0.135
	1D Soln.	0.248	0.365	0.320	0.410	0.245	0.291	0.170	0.243	0.173	0.329
Condition 4	2D Soln.	0.003	0.009	0.035	0.046	0.013	(0.000)	0.058	0.004	0.001	0.014
	1D Soln.	0.036	0.125	(0.316)	0.015	0.023	(0.293)	(0.000)	0.230	0.139	0.003

It was initially decided to take a stress value of  $7\frac{1}{2}\%$  as the criterion for a satisfactory solution. According to Kruskal's (1964) evaluation this is midway between a "good" and a "fair" fit between the raw data and the multidimensional scaling, although Klahr (1969) has challenged this evaluation. This cutoff point of 0.075 leaves us with 24 of the 30 3D solutions in conditions 1, 2 and 3 but only one of the 2D and 1D solutions as satisfactory scalings of the data. (Only two of the 3D solutions slightly exceed a stress of 10%.) The 3D solutions will be taken as the best solutions for conditions 1 to 3.

Because of the relatively small number of paired comparisons involved in condition 4, one of the 2D solutions and 4 of the 1D solutions proved to be insoluble. Table 5:2 suggests that the 2D solutions most adequately capture the underlying dimensionality for condition 4. If we increase the cutoff value to 10% stress to include borderline cases then we have available for analysis 18 3D solutions from conditions 1 and 2 combined (in which similarity of shapes alone was judged) and 10 3D solutions from condition 3, along with 9 of the 10 2D solutions from condition 4.

Two considerations are of interest to us in these analyses:

- (a) the pairs of poles which showed the widest separation on a given dimension for each solution, and
- (b) the pairs of poles which exceeded an arbitrary cutoff criterion of  $\pm 0.40$  units in the varimax rotated configurations for each solution.

The following table lists the two polar stimuli which

emerged with greatest separation on each dimension for the 40 subjects under the 4 experimental conditions:

TABLE 5:3

Condition 1

SUBJECT	1	2	3	4	5	6	7	8	9	10
Dimension 1	AC	-	AD	AF	AF	AF	BE	AC	AE	AD
" 2	BE	-	BH	CD	CD	BD	CG	AF	AH	BF
" 3	FH	-	CF	EH	CH	CG	FH	DH	CF	CE

Condition 2

SUBJECT	1	2	3	4	5	6	7	8	9	10
Dimension 1	AH	BF	AG	AD	-	AD	AH	AF	AE	AE
" 2	BE	CD	BC	BE	-	CF	BE	BH	BH	AH
" 3	CD	EH	EF	CG	-	GE	CF	CD	DG	CF

Condition 3

SUBJECT	1	2	3	4	5	6	7	8	9	10
Dimension 1	AF	AF	AE	AC	AH	AE	AE	AG	AH	AC
" 2	BC	BH	BG	DG	BG	BH	BH	BH	BD	BE
" 3	BD	CG	CH	FH	CE	CF	CG	CF	CG	BH

Condition 4

SUBJECT	1	2	3	4	5	6	7	8	9	10
Dimension 1	AC	AD	AD	AE	AD	-	AE	AB	AC	AB
" 2	CE	CD	CE	DE	BE	-	BE	CE	DE	CE

It is also useful to have information on the ways in which dimensions co-vary within individual subjects; for this we need to examine the solutions calculated for individual subjects in terms of

- (a) sets of mutually exclusive dimensions, and
- (b) sets of two or more dimensions which frequently occur together within individual subjects.

Table 5:4 gives a subject-by-subject listing of all stimuli which emerged as poles with separations greater than  $\pm 0.40$  on the Varimax rotated configurations.

TABLE 5:4

Conditions 1 and 2 (Shape Similarities).

Subject (1)	BE DE	Subject (1)	AE BE
	FH		CD CF
	CA CG		AH
Subject (2)	-	Subject (2)	GH
	-		AD CD
			BF
Subject (3)	CF	Subject (3)	AC AG
	BG BH		BC BD
	AD		E * F
Subject (4)	CD	Subject (4)	AD DH
	EG EH		BE BF
	AF		CG CH
Subject (5)	CD	Subject (5)	-
	CE		-
	GH		-



Subject (6)	AF	Subject (6)	AD DH
	BD DH		CF
	CE CG		EG
Subject (7)	BE BH DE DH	Subject (7)	AF CF
	FH		AG AH DG DH
	CG		BE
Subject (8)	DH EH	Subject (8)	AF AG
	AC AG		BH
	AF BF		CD
Subject (9)	AC CF	Subject (9)	AE AF
	AE		DG FG
	AG AH		BH CH
Subject (10)	AD*	Subject (10)	AE
	BF*		BH
	CE		CF

Condition 3 (Shape plus Label similarities).

Subject (1)	BC	Subject (4)	EG
	AE AF		FH
	BD		AC
Subject (2)	BH DH	Subject (5)	BG
	CG*		AH
	AF		CE DE
Subject (3)	AE	Subject (6)	BG BH
	CH		CF
	BG		AD AE

Subject (7)	BH DH	Subject (9)	CF CG
	AE BE		AD AF BD BF
	CF CG		AH
Subject (8)	AG	Subject (10)	BH
	BF CF		AC AE
	BH		AH

Condition 4 (Label Similarities).

Subject (1)	BC	Subject (6)	-
	AC		-
Subject (2)	BC BE CD DE	Subject (7)	B*E
	AD		AE
Subject (3)	C*E	Subject (8)	CE
	AC AD		AB AD
Subject (4)	DE	Subject (9)	D*E
	A*C*		AC
Subject (5)	BE	Subject (10)	CE
	A*D*		AB

The asterisk denotes a pole which did not achieve a separation greater than  $\pm 4.0$  in the Varimax rotated configurations, but for which no other solution was available on that dimension.

The next step requires the interpretation of the underlying structure of these extracted dimensions. First, however, a number of technical points need to be discussed. Two alternative approaches are available when we conduct multidimensional scalings of data

generated by groups of subjects under a specific experimental treatment. The first approach is to scale the responses generated by each individual subject and then pool the extracted dimensions in the attempt to come to an understanding of the underlying dimensionality of the experimental treatment task - we have used this approach here.

A second alternative is to pool data from subjects first, and then perform the multidimensional scaling. We tend to lose information and risk gaining invalid results under both approaches. Gregson (in preparation) has compared the two approaches and his study gives rise to a number of important points. By pooling the results from a series of individual analyses we can arrive at a configuration of points which is in fact representative of no individual subjective strategy. For example, suppose five subjects respond to a set of nonsense shapes in terms of a strong angularity dimension with a subsidiary closure dimension while another five subjects respond predominantly in terms of closure, with a secondary angularity dimension. By pooling we come to an overall configuration in which both closure and angularity emerge as equally dominant strategies whereas in fact no one subject responded in quite this way. It will be readily apparent that a number of such distortions can arise, including the summation of various strategies that were actually used in isolation by various individual subjects.

The alternative method, the analysis of pooled data can be even more misleading since fewer dimensions may emerge as indicated by lower stress values for a given  $n$ -dimensional solution, and as

with the first approach, the dimensions which do emerge may not be representative of individual response strategies. Thus, a considerable degree of caution is needed, no matter which approach we use. Gregson suggests that, as a compromise between the two approaches, a case could be made for grouping subjects into subgroups before performing nonmetric scalings. The above points stress the necessity for caution when we interpret the multidimensional scalings summarised here. The fact that stresses are high for the solutions obtained also suggests a need for caution in interpreting the analyses reported here.

It has been pointed out that conditions 1 and 2 are almost identical in terms of experimental treatment . thus . data from these two conditions has been pooled - it will henceforth be referred to as the "shape-similarities" group. Taken singly, stimuli A, C and F occur most frequently as the most widely separated stimuli on a given dimension as shown below (this is condensed from Table 5:3):

STIMULI	A	B	C	D	E	F	G	H
FREQUENCY	19	12	18	10	12	14	6	13

We can also tally the frequencies with which each stimulus pair emerged as

- (a) the two most widely separated poles on any dimension, and
- (b) poles which were separated by more than +4.0 units in the Varimax rotated configurations:

TABLE 5:5

	AB	AC	AD	AE	AF	AG	AH	BC	BD	BE	BF
(a)	-	2	4	3	5	1	4	1	1	5	2
(b)	-	4	4	4	5	4	3	1	2	5	4

	BG	BH	CD	CE	CF	CG	CH	DE	DF	DG	DH
(a)	-	3	5	1	5	3	1	-	-	1	1
(b)	2	5	5	3	6	4	2	2	-	2	6

	EF	EG	EH	FG	FH	GH
(a)	1	1	2	-	2	-
(b)	-	2	2	1	2	2

The interpretation of the dimensional structure underlying the 8 nonsense shapes is necessarily fairly subjective; however, a certain amount of useful information was obtained from individual subjects during informal discussion of the experiment after they had completed their judgement tasks. One strategy which was frequently described as a basis for judging shape similarities, was a tendency to respond to the "cup-shapedness" or "closure" of certain stimuli; another common strategy appeared to be based on the "angularity" or "pointedness" of the stimuli. With information such as this we can attempt to work back to the underlying perceptual structure of the shapes.

Let us first consider the "closure" dimension. This is best exemplified by stimuli D, E and F. Note that the pairs DE, DF and

EF occur very infrequently as the poles to a dimension while taken individually the stimuli D, E and F occur frequently (in total, 36 times) as one pole of a dimension when paired with some other stimulus. We would expect the opposite pole of a "closure" dimension to be exemplified by relatively "solid" stimuli with most of their area concentrated centrally, rather than peripherally as in the case of stimuli D, E and F. The "solid" shapes appear by inspection of figure 5:1 to be stimuli B, C and H. If there is a dimension of "solidity / closure" then we would expect that

- (a) there will be relatively few BC, BH and CH poles extracted as the most widely separated poles on a dimension (in fact, there are a total of 5), and
- (b) the combinations BD, BE, BG, BH, CD, CG, CE and CH would be expected to occur relatively frequently; in fact, their frequencies of occurrence are 1, 5, 0, 3, 5, 1, 3 and 1. There is thus some slight evidence for a "solidity / closure" dimension in the subjects' responses.

A second possible dimension could be described as "roundness / angularity", as exemplified by stimuli B and C (with relatively few re-entrant angles) for roundness and stimuli D, H, G and E (with many re-entrant angles) for the "angularity" pole. (Stimulus B has 5 angles which are less than 90 degrees; C has 8 such angles, while D, E, G and H have between 12 and 14 such angles, each.) In support of this postulated dimension, reference to the table above will show that

- (a) the combinations expected to occur with low frequencies, i.e.

BC on the one hand, and DE, DG, DH, EG, EH, and GH on the other occur in total only 6 times as the most widely separated poles on a dimension, and

(b) the combinations expected to occur with greater frequency, namely BD, BE, BG, BH, CD, CE, CG, and CH in fact occur with respective frequencies of 1, 5, 0, 3, 5, 1, 3, and 1 (total 19). There is thus some measure of support for a "roundness / angularity" dimension.

A third dimension suggested by some subjects during post-experimental discussions was the presence or absence of a long, thin, 'linear appendage'. This is best exemplified by stimuli A, H and to a lesser extent, C, while its opposite pole seems to be suggested by stimuli such as E, F and G. If this dimension exists as a basis for some of the subjects' responses, we would expect few solutions involving the pairs AC(2), AH(4), CH(1), EF(1), EG(1) and FG(0) as the most widely separated stimuli on a dimension; while the relatively frequently occurring combinations should be AE(3), AF(5), AG(1), CE(1), CF(5), CG(3), EH(2), FH(2) and GH(0). There thus appears to be some evidence for a weaker third dimension "appendage / no appendage".

In the above argument we have assumed that subjects judged the shapes purely in terms of physical contour. The possibility remains however, that, in the absence of experimenter-provided overt labels for the shapes, the subjects may, in fact, covertly label the shapes and then assess similarity in terms of these covert labels.

There is evidence, which will be presented later, that the associations which were most frequently given to the shapes by the subjects in the "shape-similarities" group can be classified in two dimensions:

- (1) an "animal / vegetable" dimension, and
- (2) a much stronger "meaningful / meaningless" dimension.

Now, if subjects tended to assess shape similarities in terms of the dimensions which they assigned to the stimuli when requested to overtly label the shapes, then this underlying dimensionality should emerge in the shape similarity judgements. Arguing from the responses given on the association tests by subjects from conditions 1 and 2, we would expect an "animal / vegetable" dimension to be most strongly polarised in terms of stimuli A (which is the only stimulus to which a vegetable association was given more frequently than an animal association) and B and C (both of which frequently elicited animal associations. In fact, the actual frequencies with which the pairs AB and AC emerged as the most widely separated points on a dimension were 0 and 2 respectively, thus providing very little support for the hypothesis that subjects covertly labelled the shapes then assessed similarity on an "animal / vegetable" dimension. A stronger dimension which emerged in the multidimensional scaling of the labels was a "meaningful / meaningless" dimension. The associations given to each shape in conditions 1 and 2 suggested that stimuli 1 to 4 were meaningful while stimuli 5 to 8 were frequently classified as meaningless (as shown by the number of responses of "nothing" that were given on the association



test).

If we examine the frequency with which each pair of stimuli appears as the most widely separated points on a dimension, we arrive at the following tables:

Expected low frequency occurrences:

AB(0), AC(2), AD(4), BC(1), BD(1), CD(1), EF(1),  
EG(1), EH(2), FG(0), FH(2), GH(0). (Total 15)

Expected high frequency occurrences:

AE(3), AF(4), AG(1), AH(4), BE(4), BF(2), BG(0),  
BH(2), CE(1), CF(4), CG(3), CH(1), DE(0), DF(0),  
DG(1), and DH(1) (Total 31, ~~for the same number of pairs.~~)

There is thus a fair degree of support for the hypothesis that a "meaningful / meaningless" dimension may have influenced the similarity judgements generated by subjects under the condition in which no labels were overtly paired with the stimuli. (In fact, this finding confirms the analysis reported earlier in this chapter where high association value shapes tended to be judged less similar to low association value shapes, compared with judgements between high/high and low/low pairs.) However, there appears to be no evidence that a more specific covert categorisation of classification of the stimuli by subjects influenced their judgements.

A more convincing interpretation of this point could be made by comparing the multidimensional scaling of each individual subject's

responses with the associations that he gave to each stimulus since much relevant information is lost by pooling data across subjects. If, for example, subject x gave associations of "nothing" to stimuli F, G and H but "animal" responses to stimuli B, C and D, and if his shape similarity judgements were based on his verbal classification of the stimuli, then we would expect stimuli B, C, and D to emerge at one pole and stimuli F, G, and H to emerge at the other pole of a dimension. If, however, his judgements were based on some other schema then this pattern would not emerge. In fact, only one of the twenty subjects showed a clearly marked tendency for the two classes of association responses (A = squirrel; B = elephant; C = man; D, E, F, G, H = "nothing".) to emerge as two distinctly separated clusters of points in the multidimensional scaling. The more usual tendency was for all shapes that aroused the association "nothing" to be spread throughout the space, while stimuli with similar associations, e.g. "crab", "pincer", showed no tendency to emerge as closely distributed points on any dimension.

Interpretation is necessarily subjective with multidimensional scaling but it is the feeling of this writer that the majority of subjects did judge shape similarities largely in physical terms and independently of the associations which they gave to the stimuli either before or after they judged the similarities of the shapes.

Earlier references were made to the multidimensional scalings of the labels used in condition 4. These analyses will now be discussed in detail. The stimuli consisted of 5 words; TREE, BEAR, SWAN, CRAB and NOTHING. Only 9 solutions were available

but for 8 of the 9 subjects under consideration, stimulus E (or 5; "NOTHING") emerged as one pole of one or more of the dimensions. The opposite poles are A (twice) B (twice) C (four times) and D (twice). We can thus conclude that there is a strong dimension of "animate object" / "nothing" or "meaningful / meaningless shape". Of the remaining possible dimensions, AB, AC and AD each occur twice; these pairs are TREE / BEAR, TREE / SWAN and TREE / CRAB, and taken together they suggest a dimension of "vegetable / animal". It is of interest to note that of the three remaining possible combinations, BC and BD do not occur as the most widely separated poles of a dimension, while CD (swan / crab; perhaps "vertebrate / invertebrate") occurs only once as a secondary dimension. The distribution of points in 2-dimensional space showed a striking consistency across subjects with a cluster A, B, C, D spread on one dimension and widely separated on the other dimension from stimulus E. This pattern appeared clearly in the scalings of the data for 8 of the 9 subjects that were available.

Finally, we wish to determine whether the pairing of labels with shapes induces a change in the underlying dimensionality of the similarity judgements. Several alternatives are possible, depending on the weights which subjects assign to the labels. In the extreme cases the labels may either be assigned zero weights (i.e. labels are neglected and similarity is judged in much the same way for labelled shapes as for unlabelled shapes) or alternatively the shapes may be given minimal weights and the similarity of the labels alone may dominate the judgement. It seems more likely,

however, that subjects will use a strategy somewhere between these two extremes and the question to be answered is whether the overall similarity is best described as a sum, a mean, or some other transformation of the information provided by both the shapes and labels separately. Although the multidimensional scalings for the unlabelled shape judgements show little intersubject consistency, the scalings of the label similarities were highly consistent. It should therefore be possible to show whether or not the experimenter-provided labels have influenced the shape similarities by searching for a pattern of point clusters in the labelled-shape scalings which resembles the pattern shown in the label similarity scalings. If the "animate object / nothing" dimension, which was very marked in the label scalings, has become superimposed on the 3 dimensions extracted for the unlabelled shape judgements, we would expect to find a tendency for stimuli A, B, C, and D to cluster at one pole and stimuli E, F, G, and H to cluster at the other pole to form one of the 3 dimensions. The second label-similarities dimension ("animal / vegetable" objects) is unlikely to dominate the physical shape dimensions as its effect was much weaker than the "animate object / nothing" dimension in the judgements between label similarities. In fact, by inspection of the labelled shape analyses there appeared to be little evidence for such a trend. One subject showed a clear clustering of points A, B, C, and D, while a second subject showed a clear cluster of points E, F, G, and H. Two other subjects showed a marked emergence of two clusters A, B, C, and D, E, F, G, H, suggesting that there was a strong labelling effect

for those subjects, but that stimulus D bore some similarity to the shapes E, F, G, and H, which cut across the labelling effect. The distribution of points for the remaining 6 subjects showed no apparent pattern, with high intersubject variability. The multidimensional scalings thus do not give a clear-cut description of the effects of labels on the dimensionality of the subject's responses. Much of the difficulty arises from the degree of subjectivity that is required to interpret the scaling solutions. However, the complexity of the experimental situation, particularly with regard to the fact that the effect of labels is definitely second order in comparison with the contribution of the actual physical stimulus differences themselves is also an important factor. It appears likely that actual strategies used by subjects in combining the information that is present in both the shapes and the labels is highly variable, so that some subjects may give minimal weight to the labels while others use them as the main basis for categorising or classifying the stimuli.

In order to gain a better understanding of the way in which the overall stimulus complex shape-plus-label similarity is judged relative to the two components shape similarity and label similarity, it was decided to extend the model which has been sketched in preceding chapters.

Since the overtly labelled shape carries more information than either the shape alone or the label alone, it was suggested that a model which could be applied to the integration of information in judgement processes would be applicable to the problem of

describing how the two components, physical shape and overt label, are combined into an overall assessment of similarity. One such model has been developed by Anderson (1970 ). In very general form it predicts that when a subject is required to reach some kind of judgement on a set of stimuli, the judgement will be a linear function of the values of the items:

$$J = C + \sum w_k S_k$$

where  $J$  is the judgement,  $C$  is a constant and  $S_k$  and  $w_k$  are the scale value and weight of the  $k$ th item in the set.

When the items are presented sequentially, as in attitude change experiments, it can be assumed at least in the simplest applications, that each new piece of information will be integrated with the then current judgement to yield a revised judgement. In this special case, the model may be written,

$$J_k = C_k S_k + (1 - C_k) J_{k-1}$$

where  $J_k$  and  $J_{k-1}$  are the values of the judgement after and before presentation of the  $k$ th stimulus,  $S_k$  is the scale value of the  $k$ th stimulus and  $C_k$  is the change parameter, (Anderson, 1968). Mathematically, this is an averaging, rather than an adding model; the more general model however, although formally an adding model, includes averaging subtracting, dividing and multiplying models, as well as proportional change models (Anderson, 1970 ).

In the present experiment we have the situation in which the similarities of labelled shapes, unlabelled shapes and labels have been judged by independent groups of subjects. Ideally the Anderson model should be applied to judgements made by single

subjects who assessed individual similarities under all 3 conditions. As an initial investigation however, it was decided to attempt to apply the model by testing individual subjects' judgements of labelled shape similarities against the mean judgements given by the 10 subjects in each of the shape similarities and the label similarities conditions.

Let A represent the similarity between a pair of labelled shapes

" B " " " " " " " shapes

" C " " " " " " " labels

From Anderson's model, for the case of sequentially presented items, we can postulate that for a given individual  $i$  and a given pair of stimuli  $j, k$  the similarity between labelled stimuli will be given by

$$A'_{ijk} = w_i B_{jk} + (1 - w_i) C_{jk}$$

where  $w$  is a weight which corresponds to the constant  $C$  in Anderson's equations and  $A'$  is the calculated or predicted value of  $A$ . Note that when  $w = 1$ , labelled shape similarities are equal to unlabelled shape similarities, i.e. the labels have no effect on the judgement of  $A$ . Conversely, when  $w = 0.0$  the labelled shape similarities are equal to the label similarities, i.e. subjects are judging entirely in terms of the names that we have paired with the shapes.

A priori, we expect  $w$  to be somewhere between 0 and 1, it is of theoretical interest to know just what weight is assigned to the labels and how much the weight varies from subject to subject. Anderson (1968 , 1970) has stressed the importance of measuring

the goodness of fit of a model to the data it is considered to represent; he has extensively used analysis of variance as a measure of goodness of fit.

For the present experiment, however, it was decided to measure goodness of fit in terms of percentage error. The best solution is the one which minimises the sum of the squared differences between the predicted values of  $A'_{ijk}$  (as derived by solving the equation in terms of  $B_{jk}$  and  $C_{jk}$ ) and the experimentally observed values of  $A_{ijk}$ . The squared differences are divided by the sum of the squared values of  $A_{ijk}$  (experimentally measured) to yield the percentage error:

$$\text{PERCENT ERROR} = \frac{\sum (A_{ijk} - A'_{ijk})^2}{\sum (A_{ijk})^2}$$

The minimum value of the percentage error, with its associated  $w$  (and  $v$ ) values gives the best solution for the model.

Gregson (personal communication) has suggested two extensions of the Anderson model which made allowance for the fact that the judgements made by subjects frequently appear to exhibit adding, rather than averaging, under certain conditions.

The data were analysed in terms of 3 models; model 1 is the averaging model of Anderson, given above.

Model 2 is as follows:

$$A'_{ijk} = [w_i B_{jk} + (1 - w_i) C_{jk}] \left[ \frac{1}{2} \frac{(B_{jk} - C_{jk})^2}{(B_{jk} + C_{jk})} \right]$$

The exponent to which the multiplier  $\frac{1}{2}$  is raised has the effect of



reducing the value of  $A'_{ijk}$ . The reduction is slight when  $B_{jk}$  and  $C_{jk}$  differ only slightly in scale value, however when the judgements produced between pairs of shapes ( $B_{jk}$ ) and pairs of labels ( $C_{jk}$ ) are widely separated along the scale the exponent can reduce the magnitude of the predicted  $A'_{ijk}$  quite considerably. This modification captures the situation in which adding occurs when the magnitude of separation in scale value for  $B_{jk}$  and  $C_{jk}$  is large, but averaging occurs when  $B_{jk}$  and  $C_{jk}$  are close in scale value.

Gregson also suggested a third model which took the following form:

$$A_{ijk} = [w_i B_{jk} + (1 - w_i) C_{jk}] \left[ \frac{1}{2} \frac{v|B_{jk} - C_{jk}|}{(B_{jk} + C_{jk})} \right]$$

In this case, the modulus or absolute difference of the scale values of  $B_{jk}$  and  $C_{jk}$  divided by the sum of the scale values is used. In itself, this gives an exponent which is always less than 1.0, but by including an additional multiplier,  $v$ , it is possible for the exponent to vary between zero and any positive value we choose. In the present experiment,  $v$  was varied between 0.2 and 2.0 in increments of 0.2; this gives a maximum magnitude to the multiplier of  $0.5^{1.64}$  (or 0.322) which occurs in the case where  $v = 2.0$  and  $B_{jk}$  and  $C_{jk}$  take the maximum possible difference in scale values. This multiplier again has the effect of lowering the calculated value of  $A'_{ijk}$ , however model 3, unlike model 2 allows for the possibility that adding may occur not only when  $B_{jk}$  and  $C_{jk}$  differ widely in scale value, but also when they are

close in scale value.

Note that as we have assigned a scale value of 1.0 to the "identical" pole of the similarity scale, then adding occurs when  $A_{ijk}$  is less than both  $B_{jk}$  and  $C_{jk}$ , averaging occurs when  $A_{ijk}$  is intermediate between  $B_{jk}$  and  $C_{jk}$  and subtracting occurs when  $A_{ijk}$  is greater than both  $B_{jk}$  and  $C_{jk}$ . None of the models used in this experiment allows for subtracting since by taking the square of the difference between  $B_{jk}$  and  $C_{jk}$  in one case, and the modulus of the difference, with an associated  $v$ -value of 2.0 or less in the other, we ensure that the exponent is always positive.

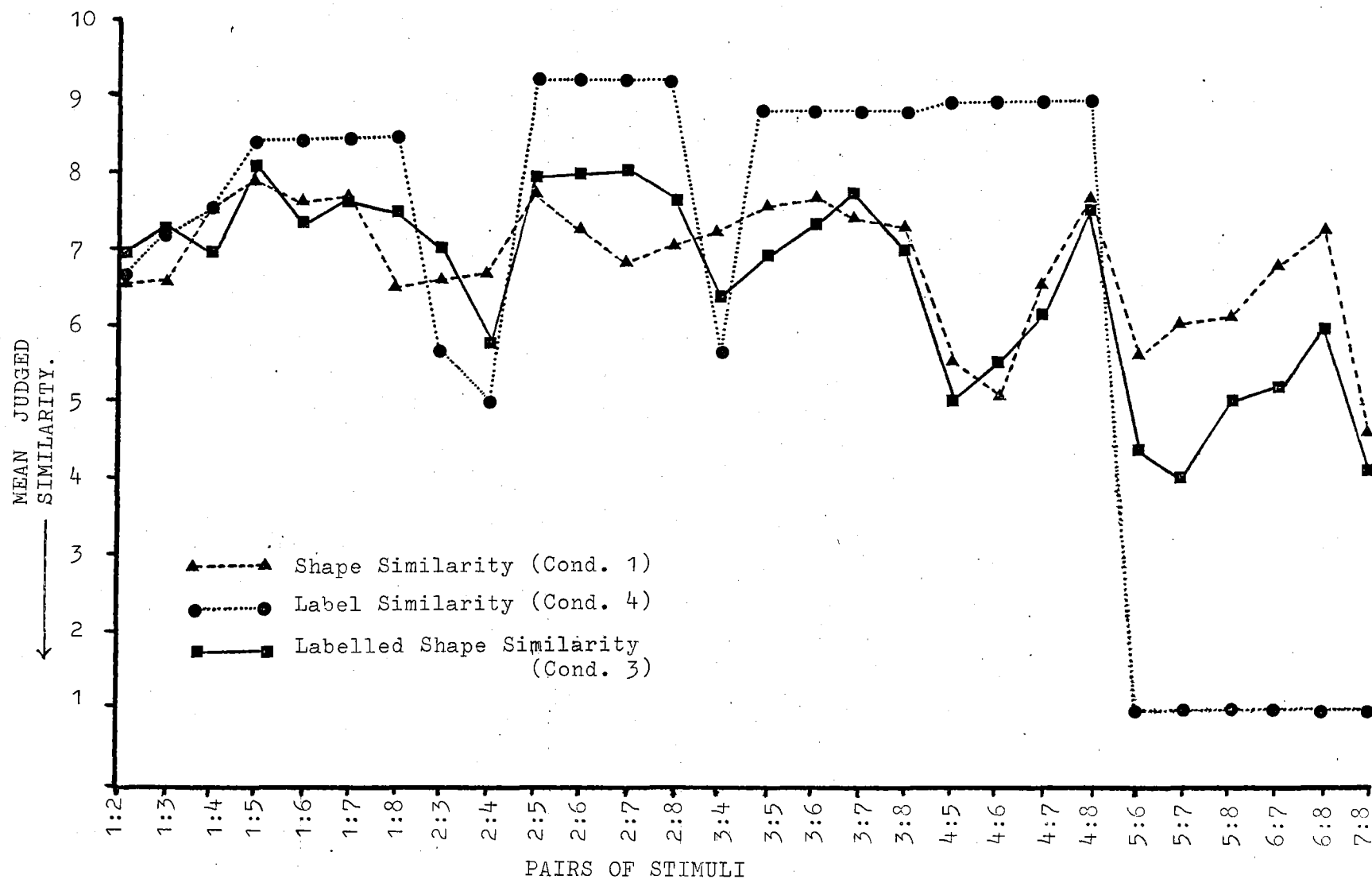
The mean judgements made by each of the 10 subjects in conditions 1, 3 and 4 are shown as pooled data in figure 5:2. Note the subtracting occurs at only 3 points but adding occurs at several points where the differences between B and C are both small and quite large. That is, figure 5:2 suggests that when averaging does not occur, adding is the more frequently occurring form of the two remaining strategies. However, this graph is based on pooled data; for individual subjects the pattern of responses may be quite different.

A computer program was written to calculate the goodness of fit, as measured by percentage error for models 1, 2 and 3 under  $w$  values ranging from 0.0 to 1.0 in increments of 0.05. For model 3 the values of  $v$  ranged from 0.2 to 2.0 in increments of 0.2.

Table 5:6 lists the minimum percentage errors and the corresponding  $w$  (and  $v$ ) values for the three models. The individual responses made by each of the 10 subjects in condition 3 of

FIG 5:2

POOLED DATA FROM EXPERIMENT 4 - THE EFFECT OF LABELS  
ON JUDGED SIMILARITY OF NONSENSE SHAPES.



experiment 4 (i.e. the  $A_{ijk}$  values for subjects 1 to 10) are evaluated against the pooled mean responses made by the 10 subjects in condition 1 ( the  $B_{jk}$  values ) and the 10 subjects in condition 4 (the  $C_{jk}$  values).

TABLE 5:6     MINIMUM PERCENTAGE ERRORS AND ASSOCIATED  
W (AND V) VALUES FOR SUBJECTS 1 TO 10 TESTED UNDER MODELS 1, 2, & 3.

The asterisk denotes the best solution for a given subject.

MODEL 1:

Subject	Min. % Error	W Value
1	4.42*	0.15
2	3.18*	0.55
3	9.02	0.20
4	5.76*	0.00
5	9.39	0.25
6	3.68	0.15
7	8.85	0.35
8	4.69	0.20
9	3.61	0.20
10	4.60*	0.05

MODEL 2:

Subject	Min. % Error	W Value
1	18.55	0.65
2	9.87	1.00
3	15.20	0.00
4	23.82	0.65
5	15.03	0.20
6	12.55	0.25
7	11.76	0.55
8	8.28	0.15
9	12.99	0.45
10	18.56	0.55

MODEL 3:

Subject	Min. % Error	W Value	V Value
1	4.98	0.05	0.20
2	3.53	0.55	0.20
3	6.71*	0.00	0.80
4	6.57	0.00	0.20
5	6.43*	0.00	1.20
6	2.85*	0.00	0.60
7	6.41*	0.20	1.00
8	2.10*	0.00	0.80
9	3.49*	0.15	0.20
10	4.96	0.00	0.20

Model 3 provides a better fit with the data for 6 of the 10 subjects; the remaining 4 subjects favour model 1. Note that  $v$  took on its minimum possible value of 0.2 in all cases in which model 1 produced a better fit than model 3; this is to be expected since, as  $v$  tends to zero the multiplier in model 3 tends to unity, and model 3 then becomes identical to model 1.

In all except one case, the  $w_i$  values associated with the best fit tend to be small for both models 1 and 3. This suggests that the judgements have been almost entirely dominated by the labels, however this may be attributed to the fact that we have compared individual judgements of  $A_{ijk}$  with the pooled values of  $B_{jk}$  and  $C_{jk}$ . Since estimates of shape similarity show wider inter-subject variability than estimates of label similarity, the pooling of data will have produced a greater regression towards the mid-point of the scale for the pooled shape judgements than for the pooled label judgements. The pooled mean responses to the labels will hence show a comparatively wider range of scale usage, as will the individual estimates of  $A_{ijk}$ . The  $C_{jk}$  and the  $A_{ijk}$  pairs of corresponding points will both tend to be separated from the  $B_{jk}$  values, thus producing a spuriously low weighting of the  $B_{jk}$  (shape similarity) values in the models.

Model 2 does not provide a good representation of the process involved - this can be explained in terms of the points noted earlier: Model 2 predicts that

- (a) adding occurs when  $B_{jk}$  and  $C_{jk}$  are widely separated, and
- (b) averaging occurs when  $B_{jk}$  and  $C_{jk}$  are close in scale value.

In fact, the data suggest the opposite trend; averaging occurs when  $B_{jk}$  and  $C_{jk}$  are widely separated but adding (or perhaps to a larger extent, subtracting) occurs when  $B_{jk}$  and  $C_{jk}$  are close in scale value. It is thus not surprising that the percentage errors are high for model 2. The best solutions for models 1 and 3 tend to have high error terms (a value of less than 5% error seems to suggest a reasonably good fit between data and model); however, this is not surprising when we consider that individual judgements of labelled shape similarities have been tested against pooled shape similarity and label similarity judgements from independent groups of 10 subjects. It was therefore decided to perform a final experiment in which subjects each judged under all 3 of the experimental conditions:

- (a) judgements between unlabelled shapes,
- (b) judgements between labelled shapes, and
- (c) judgements of the similarity in meaning of the labels.

This experiment will now be described in detail.

#### EXPERIMENT 5.

The stimuli used were the 8 Vanderplas and Garvin (1959) nonsense shapes as used in the preceding experiment. The label "nothing" was omitted from the set however, to give the following label-shape pairings:

Stimulus	Label
1	TREE
2	BEAR

3	SWAN
4	CRAB
5	TREE
6	BEAR
7	SWAN
8	CRAB

The subjects were 20 volunteer graduate and undergraduate students. Each subject was paid \$1.00 for participating; the experiment took less than one hour per subject. To balance for temporal factors in the judgement processes, ten subjects judged shapes first followed by labels, and finally the labelled shapes, while the remaining ten subjects judged labelled shapes first, followed by labels and finally the unlabelled shapes. As in experiment 4, a complete method of paired comparisons was used in all 3 parts of the experiment. The 8 stimuli gave 56 judgements in conditions 1 and 3; identical pairs of stimuli were not presented. The actual order of presentation of stimuli was random, and varied cyclically from subject to subject. For each individual subject the labelled shapes were presented in a different order from the unlabelled shapes. Subjects were permitted to examine the stimuli ad. lib. before they were required to assess similarities.

The instructions to the first 10 subjects are shown below; they were suitably modified for subjects 11 to 20 for whom labelled shapes rather than unlabelled shapes were judged in Part I.

#### "INSTRUCTIONS TO SUBJECTS.

There are three parts to this experiment - all three parts



involve the making of similarity judgements between pairs of stimuli. You are to make judgements using the scale below:

1. Identical
2. Nearly identical
3. Very similar
4. Similar
5. Slightly similar
6. Slightly dissimilar
7. Dissimilar
8. Very dissimilar
9. Extremely dissimilar
10. Completely different.

When you make a judgement between a pair of stimuli, simply say to me the number that is paired with the appropriate word or phrase used in the ten-point scale above. You may keep these instructions beside you.

#### PART 1.

In part 1 you will be judging the similarity between pairs of unfamiliar shapes called "nonsense shapes". Most people see some of the shapes as representing all kinds of familiar objects. However, please try to ignore any objects or situations that each shape may remind you of, and try to judge similarity in terms of the actual physical contour or outline of the shape alone. There are a total of 56 pairs if we judge each of the 8 shapes in turn against every other shape. Before you make your judgements, I will show you the whole set of 8 shapes, so that you can get some idea of how much the

shapes vary from one to another.

## PART 2.

In part two, you are required to judge the similarity in MEANING of a number of words. You will use the same ten-point response scale as before. Try to respond in terms of the general meaning of the words as such, do NOT try to respond to particular examples of the objects or situations described by each word. In particular, you must avoid, for the present, any tendency to refer the words back to the shapes that you have seen earlier in this experiment. Here are the four words that will be used:

TREE, BEAR, SWAN, CRAB.

## PART 3.

Part 3 combines the stimulus material used in parts 1 & 2. This time you are required to attempt to integrate the information present in both the shapes and the labels, to produce an OVERALL assessment of similarity. That is, when judging each labelled stimulus, try to take account of both the physical shape and the meaning expressed by the label to come to an overall assessment of the similarity of the total stimulus complex."

Note that the instructions are designed to maximise the shift in judgements produced by labelling in that subjects are requested to judge shape similarity on the basis of physical contour alone, while the judgements between labelled shapes require both shape and label to be integrated into the overall judgement. Care was taken to ensure that neither the averaging nor the adding / subtracting

strategy was suggested in the instructions; in fact, one of the aims of this experiment is to determine which of the strategies the subjects will use spontaneously.

## RESULTS AND ANALYSES.

Symmetrically opposite pairs of stimuli were combined for each subject and their means were calculated to give two 8 x 8 half-matrices of unlabelled and labelled shape similarities and a 4 x 4 half-matrix of label similarities.

The computer program for models 1, 2 and 3 was altered to calculate the goodness of fit of the data when independent measures of  $A_{ijk}$ ,  $B_{ijk}$  and  $C_{ijk}$  are available for each individual subject. The mathematical structure of the models themselves, and the range of  $w_i$  and  $v$  values used was not altered. The results for each subject as tested under the three models are given in Table 5:7.

The asterisks in the percentage error columns signify the best solution for each subject. In those cases for which the percentage error as shown in the table was equal for both models 1 and 3, the data actually favoured model 1 - as shown by slightly lower minimum values of  $\Sigma(A_{ijk} - A'_{ijk})^2$ ; the differences in these values for models 1 and 3 were lost however in the rounding of numbers during division by  $\Sigma(A_{ijk})^2$ .

Although the percentage errors for model 2 are all below 1.25%, representing a good fit between the model and the data, better fits are obtained for models 1 and 3. Fourteen of the

TABLE 5:7 MINIMUM PERCENTAGE ERRORS AND ASSOCIATED  
VALUES OF W (AND V) FOR SUBJECTS 1 TO 20 TESTED UNDER MODELS 1, 2 & 3.

MODEL 1:

Subject	Min. % Error	W Value
1	0.04*	0.75
2	0.08*	0.45
3	0.18*	0.80
4	0.18*	0.70
5	0.10*	0.35
6	0.18*	0.85
7	0.53	0.20
8	0.47*	0.65
9	0.33	0.70
10	0.04*	0.50
11	0.13	0.25
12	0.21	0.50
13	0.26	0.55
14	0.06*	0.45
15	0.14	0.60
16	0.29*	0.75
17	0.07*	0.60
18	0.02*	0.25
19	0.28*	0.35
20	0.51*	0.50

MODEL 2:

Subject	Min. % Error	W Value
1	0.25	0.65
2	0.67	0.55
3	0.50	0.00
4	0.39	0.60
5	0.59	0.50
6	1.03	1.00
7	1.11	0.25
8	1.22	0.00
9	0.70	1.00
10	0.25	0.00
11	0.79	0.00
12	0.46	0.35
13	0.78	1.00
14	0.34	0.40
15	0.81	1.00
16	0.90	0.80
17	0.63	0.40
18	0.42	0.00
19	0.89	0.00
20	1.23	0.45

MODEL 3:

Subject	Min. % Error	W Value	V Value
1	0.04	0.70	0.20
2	0.09	0.40	0.20
3	0.18	0.75	0.20
4	0.19	0.65	0.20
5	0.12	0.30	0.20
6	0.19	0.85	0.20
7	0.46*	0.10	0.80
8	0.48	0.60	0.20
9	0.23*	0.75	1.80
10	0.05	0.45	0.20
11	0.08*	0.45	0.40
12	0.20*	0.40	0.60
13	0.14*	0.60	1.40
14	0.07	0.40	0.20
15	0.08*	0.65	1.20
16	0.30	0.75	0.20
17	0.08	0.60	0.20
18	0.02	0.15	0.20
19	0.31	0.30	0.20
20	0.52	0.50	0.20

subjects showed a best fit with model 1, the remaining 6 favoured model 3. Reference to the data generated by individual subjects indicates that the three processes of adding, subtracting and averaging appear to occur in all subjects but to different degrees in different individuals. The  $w$  values associated with the best fit for each subject also appear to be highly variable across subjects; however, they are approximately normally distributed with a mean of  $w = 0.54$  suggesting that in general both shapes and labels have been given approximately equal weightings in this particular experiment. We would expect however that instructions to subjects would have a large effect on the weight given to the labels; in experiment 4, the mean weight for the 10 subjects was  $w = 0.11$  which suggests that the labels were given much greater weight than the shapes in the judgement process in that particular study, although it contained a methodological artefact.

The problem of why subjects sometimes add, sometimes average, and sometimes subtract the judged similarities of shapes and of labels when coming to an overall assessment of labelled-shapes similarity needs closer examination.

Table 5:8 lists for each stimulus pair the frequencies with which the judgements of labelled shape similarity showed adding, subtracting or averaging by the twenty subjects who responded to each pair in experiment 5.

TABLE 5:8

Stimulus Pair (Stim.)	1:2	1:3	1:4	1:5	1:6	1:7	1:8	2:3
Freq. of Adding (+)	5	8	7	0	8	6	3	4
Freq. of Subtracting (-)	5	3	3	3	2	3	11	2
Freq. of Averaging (Av)	10	8	10	17	10	11	6	14

(Stim.)	2:4	2:5	2:6	2:7	2:8	3:4	3:5	3:6	3:7	3:8	4:5
(+)	9	5	0	5	5	5	4	2	0	6	0
(-)	2	4	0	7	6	3	5	4	2	2	10
(Av.)	9	11	20	8	9	12	11	14	18	12	10

(Stim.)	4:6	4:7	4:8	5:6	5:7	5:8	6:7	6:8	7:8
(+)	0	5	0	0	4	3	2	8	1
(-)	8	0	0	11	8	6	7	3	5
(Av.)	12	15	20	9	8	11	11	9	14

The results are slightly biased in favour of the averaging formulation since on those occasions in which the judgement of labelled shape similarity was given the same scale value as either the shape similarity or the label similarity the judgement was tallied under the averaging formulation.

Note that the identically labelled shapes (1:5, "tree"; 2:6, "bear"; 3:7 "swan" and 4:8, "crab" almost exclusively produce an averaging response. This suggests that when label similarities are much greater than shape similarities, averaging occurs. (Totals for the 4 pairs are zero adding, 5 subtracting and 75 averaging



responses). A priori we might also expect averaging to occur when shape similarities are much greater than label similarities, however no data is available to support this prediction since there were apparently no highly dissimilar pairs of labels and no highly similar pairs of shapes in the stimulus sets used. The data also suggest that when label similarities are only slightly greater than shape similarities, as shown by the pairs 1:4, 1:6, 1:7, 2:3, 2:4, 2:5, 2:7, 3:6, 3:8, 6:7, and 6:8, then averaging responses predominate but adding and subtracting of similarities also occur relatively frequently. Likewise, we would expect a similar trend to emerge when shape similarities are only slightly greater than label similarities as shown by the pairs 1:8, 4:5, 4:6, 5:6, and 7:8. The mean number of adding, subtracting and averaging responses per stimulus pair is shown for each of these conditions in the table below:

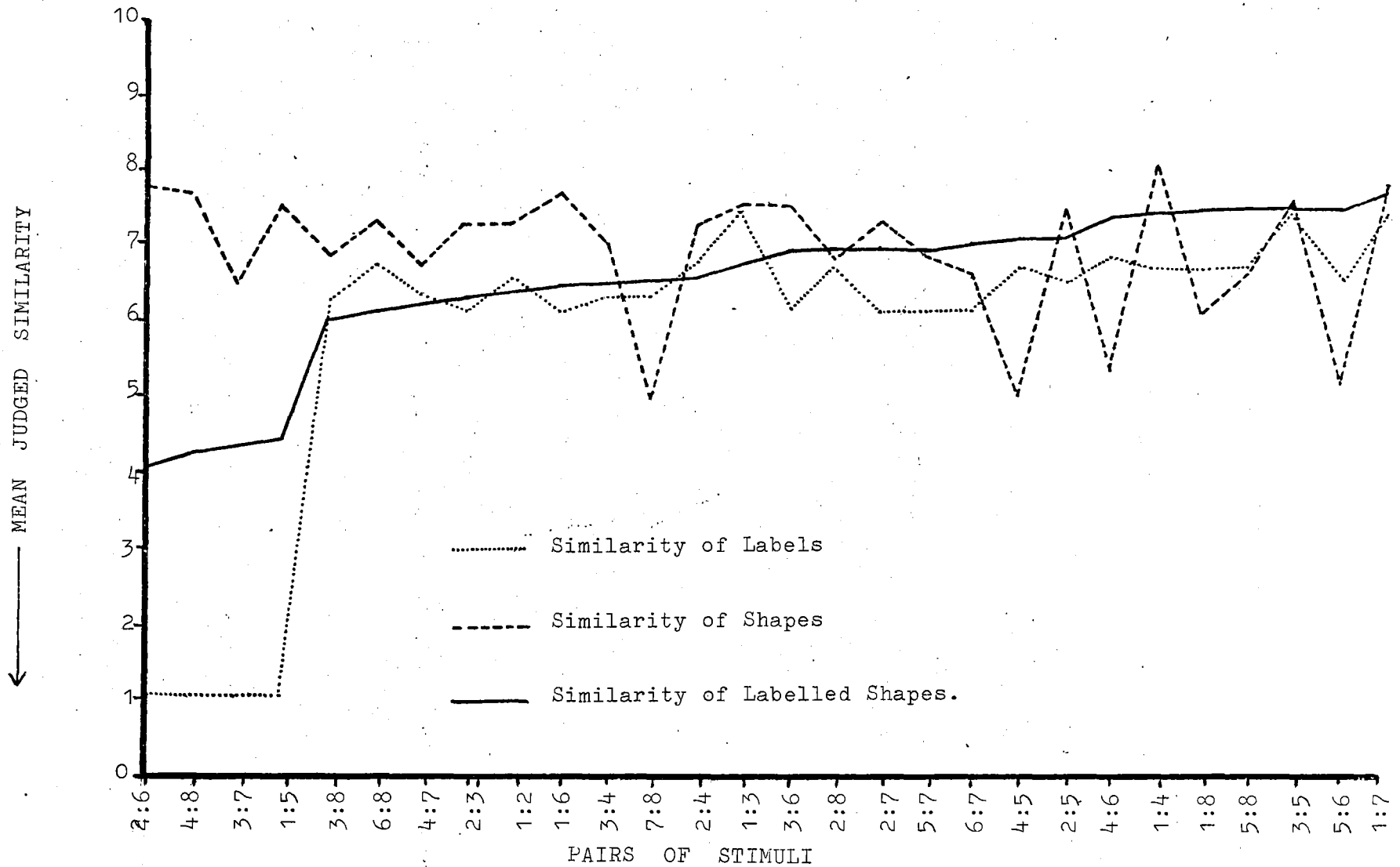
	Adding	Subtracting	Averaging
For label s much greater than shape s	0	1.25	18.75
For shape s much greater than label s	No data available.		
For label s slightly greater than shape s	5.64	3.54	10.82
For shape s slightly greater than label s	1.33	8.83	9.98

The pooled data thus suggest that subjects may be making use of three different strategies in this experiment. Averaging seems to occur when label similarity and shape similarity are widely separated in scale value. When the shape similarity and label similarity are close in scale value however, adding and subtracting strategies become more prevalent. More specifically, when label

similarity is slightly greater than shape similarity then labelled shape similarity tends to be greater than unlabelled shape similarity, i.e. either averaging or adding occurs. Conversely when label similarity is less than shape similarity, then labelled shape similarity tends to be less than unlabelled shape similarity, i.e. averaging or subtracting occurs. Figure 5:3 shows the pooled mean responses made by the 20 subjects in experiment 5. The trend towards an averaging / adding strategy when label similarity is greater than shape similarity, but an averaging / subtracting strategy when label similarity is less than shape similarity is clearly discernible. The graph also provides some evidence for a weaker trend in which adding occurs when label similarity is only slightly greater than shape similarity and subtracting occurs when label similarity is only slightly less than shape similarity. From the psychological viewpoint it seems reasonable that these trends should emerge.

When shape similarities and label similarities differ in scale value and a subject is required to combine the two dimensions into an overall judgement of similarity then the averaging approach would seem to provide a straight forward and satisfactory strategy. However, when shape similarity and label similarity are very close, or even identical in scale value, a different approach would appear to be necessary. Suppose that a subject is presented with two pairs of stimuli,  $x, y$  and  $x', y'$ ; his task is to judge the similarity of each pair; let us further assume that he assigns a scale value of 5.0 to each pair. We now present the subject with two complex

FIG. 5:3 EXPERIMENT 5 - POOLED DATA FOR  
THE 20 SUBJECTS.



stimuli  $x + x'$  and  $y + y'$  and require him to assess their similarity. On the averaging strategy he will assign a value of 5.0 to the pair of stimuli. If however he decides that the similarity of a pair of stimuli relates to the number of elements that they have in common then he may judge the bidimensional stimuli as having either a smaller or a greater number of common elements than either of  $x$ ,  $y$  or  $x'$ ,  $y'$  taken separately. In order to express this perceived shift in similarity when the dimensionality of the stimuli is increased, the subject must abandon the averaging strategy and adopt a new adding or subtracting strategy to accommodate his change in judgement.

H8ijer (1969) in a similar investigation in which unidimensional stimuli were combined to give bidimensional stimuli found that the judgements produced could be described by an additive model, however his stimuli formed a more homogeneous set than those used here, so that the averaging model ( which perhaps fits better when stimulus differences are large) may not have been applicable in his case.

Anderson (1965 a) has pointed out that neither the adding nor the averaging formulation of information integration may be entirely adequate since both are supported in the literature. In the majority of cases for which the averaging model has been applied to judgement data (e.g. Anderson, 1965 b, Anderson, 1966) the scale values of the stimuli used have been fairly widely separated in magnitude. An exception is the study reported by Anderson (1967) in which sets of adjectives were of equal value were presented

to subjects who were required to judge the likeability of the person described by the adjectives. Contrary to our present view which states that adding or subtracting occur when stimuli are of equal scale value, Anderson reports that the results could be accounted for by the hypothesis that the response is a weighted average of the adjective values and a neutral impression. Some slight support for our position is given by an experiment conducted by Anderson and Jacobson (1965). They required their subjects to judge the likableness of persons described by sets of 3 adjectives under one of 4 instruction conditions, including a condition in which each adjective was required to be rated as equally important. In this case it was reported that the simple averaging model worked reasonably well, though not perfectly. It may be then, that adding or subtracting strategies tend to dominate the judgement process when the stimuli that are to be combined in an overall judgement have similar or equal scale values, however the point needs further detailed investigation and experimental testing. The argument would be more convincing if we could show that the effect is observable not only within the pooled responses from the 20 subjects of experiment 5, but also within the sets of responses generated by individual subjects. Let us take as our hypothesis the statement that when shape similarity and label similarity differ by a scale value of 1.5 or less, then adding and subtracting responses will predominate and that furthermore, adding will predominate when label similarity is greater than shape similarity, but subtracting will tend to occur when shape similarity is greater

than label similarity. We also predict that averaging responses will predominate when shape similarities and label similarities differ in scale value by more than 1.5. For each individual subject the number of adding, subtracting and averaging responses made when the shape and label similarity scale values (a) differed by 1.5 units or less, and (b) differed by more than 1.5 units were tallied. (On those occasions in which shape similarity equalled label similarity in scale value, the responses were not included in the tally.) From the sums of these tallies, the mean number of adding, subtracting and averaging responses per subject was calculated. These means are shown in the table below:

	Small differences ( $\leq 1.5$ ) in shape and label similarities.			Large differences ( $> 1.5$ ) in shape and label similarities.		
	Adding, subtracting, averaging.			Adding, subtracting, averaging.		
Shape similarity > label similarity	1.10	1.55	1.55	0.55	0.75	3.90
Label similarity > shape similarity	1.70	1.65	2.05	0.90	1.50	8.45

There is thus a weak trend in the predicted direction, however it is masked by the high frequency of averaging responses in the two conditions where subtracting (Cell 1) and adding (Cell 3) are expected to occur. As predicted, averaging occurs more frequently when stimulus differences are large (Cells 2 and 4 of the table above).

In general then, the basic strategy for integrating the information that is present in the shapes and the labels follows the averaging formulation developed by Anderson. There are suggestions that research into the hypothesis that adding and subtracting occur when differences in scale value of the labels and the shapes are small, may prove to be fruitful.

Although our hypothesis appears to be supported to some extent by the data, it should be pointed out that the hypothesis, as stated, accounts for the majority of logical alternative trends that it is in fact possible for the data to exhibit.

More specifically, if A, B, and C refer to labelled shape, unlabelled shape and label similarities respectively, then our hypothesis predicts that

(1) If  $C > B$

Then A can be greater than B but less than C (Averaging formulation)

or A can be greater than B and greater than C (Adding formulation).

But A cannot be less than B.

(2) If  $B > C$

Then A can be greater than C but less than B (Averaging formulation)

or A can be less than C and less than B (Subtracting formulation)

But A cannot be less than C.

One way to test this hypothesis would be to alter model 3 so that it allows not only adding (as at present) but also subtracting to occur when the difference between B and C is small. This could be effected by using a larger range of  $v$  values, as suggested earlier. If this modification then improved the goodness of fit to such an extent that model 3 became superior to model 1, then this could be interpreted as favouring an hypothesis that subjects change from an averaging to an adding or subtracting strategy when differences in scale value between B and C are small. The model fitting exercise could be repeated on data for which the differences between B and C were known to be small and the point at which the hypothesised shift in strategy occurs could be determined in terms of the magnitude of difference between A and B.

These points, along with many of the other questions that have been raised by the 5 experiments described in this thesis may be answered in future research. In the chapter which follows, the main results reported to date will be summarised and the relevance of this research to the experiments described in chapter 1 will be outlined briefly.



## CHAPTER VI

### A BRIEF SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH.

In this final chapter, we will summarise the main findings and attempt to draw some general predictions concerning the role of labels in cognitive behaviour.

The literature summarised in Chapter I was reviewed in two distinct sections, neither of which included research that could be considered to bear much resemblance to the experiment approach used in the studies reported here. There are a number of reasons for this: Firstly, unlike the majority of studies in psycholinguistics we have used the judgement of subjective similarity as our approach to cognition; the present approach therefore owes more to psychophysics than to linguistics. Secondly, we have treated similarity assessment as a form of cognitive categorising behaviour, in keeping with the fourth definition given by Wallach (1958). By contrast, the majority of studies reviewed in the section on mediation, have treated similarity, if at all, in terms of primary stimulation gradients. In these cases, similarity is considered to be based on generalization and discrimination learning (the third definition given by Wallach, 1958). Other studies have investigated the role of mediation in the acquisition of relatively simple motor skills rather than in complex cognitive processes.

It was suggested that many of the mediation studies reported in Chapter I point to the need for a more adequate conceptualisation

of the ways in which subjects process information in a variety of experimental situations. Contemporary cognitive theory appears to provide for such a conceptualisation. A number of diverse approaches are currently developing which, instead of emphasising the role of conditioning in cognitive activities, draw our attention to the role of strategies, and such factors as conceptual, schematic and classificatory processes in cognition.

Experiments 1 and 2 investigated the effect of nonsense labels on judged similarity between pairs of stimuli. The results for experiment 1 suggested that the nonsense labels increased the psychological distances between similar stimuli but reduced the distance between pairs of dissimilar stimuli by producing a regression of judged similarities and dissimilarities towards a mid-scale baseline. This result was interpreted in terms of a trace decay theory which postulates that the rate of decay of memory traces is altered by pairing verbal labels with stimuli. In experiment 1, the nonsense labels were considered to increase the rate of trace decay by increasing the confusion between tastes which were paired with a set of unfamiliar and presumably meaningless nonsense words.

Later experiments however suggest that trace decay theory is not in itself sufficient to account for the shifts in judgements that are induced by labelling, for even when labelled stimuli are presented simultaneously for judgement (thus allowing no opportunity for trace decay to occur) the labelling effect still appears.

In experiment 2, the two nonsense labels used produced an

enhancement of judged similarity (or reduction in psychological distance) regardless of whether the stimuli themselves (in unlabelled form) were judged near the similar or the dissimilar end of the continuum. We have noted for experiment 1, however, that the labels increased the psychological distance between similar stimuli but decreased the distance between stimuli at the dissimilar end of the continuum. The trace decay model provides an explanation for the direction of shift in experiment 1, however in experiment 2 a different approach is required to explain the consistent enhancement of similarity for all stimuli induced by the nonsense labels. The model suggested was that the labels act as a secondary dimension such that the labelled stimuli were judged as more similar to a labelled standard square than when the same shapes were judged without labels. It was necessary to assume that the two labels used, ZIQ and QAP were functionally equivalent as elements in a dimension of similarity. It is not clear how the data from experiment 1 could be explained in terms of the dimensional model however. If the four labels used, ZUM, TOV, JEG and DAX, add a perceptually equivalent cue to the tastes, we would expect an enhancement of similarities to occur across all scale values; likewise if the labels added an extra dimension of dissimilarity we would expect a decrease in judged similarity across all scale values. Instead however, we have the two-way shift in judgements (relative to the baseline) that is reported above. It is therefore necessary to assume both the trace decay model and the dimensionality model operate under the two different conditions that were experimentally

tested, although the possibility remains that the direction of shift is related to differences in the psychophysical method employed (four standards in experiment 1, but only one standard in experiment 2).

Experiment 3 further investigated the labelling phenomenon by examining the effects of qualitative descriptive labels on similarity judgements. A comparison of two different experimental approaches, the traditional method of paired comparisons (with only brief delay between presentation of standards and comparison stimuli), and the "memorised standards" method of experiment 1 showed that when the delay between presentation of standards and comparison stimuli is long, then judgements tend to regress towards the mid-scale baseline, a finding which supports the trace decay model. When labels which adequately conceptualise the taste qualities present in the stimuli are paired with the tastes, then the degree of regression to the baseline is reduced, again supporting the decay model with its assumption that adequate descriptive labels reduce the rate of short term trace decay.

By performing a series of nonmetric multidimensional scalings of the data, an attempt was made to show that the labels add to the dimensionality of the situation. However, the actual number of dimensions that was extracted appeared to be largely unaffected by the labels, although interpretation of the results of the scalings was necessarily subjective. There was a tendency for the constant underlying dimensionality to emerge with greater or stronger separation as the precision with the labels describing the taste

qualities was increased, lending some support to the interpretation that the labels assimilate along the already existing dimensions rather than adding new dimensions to the underlying structure of the judgement task. This finding is again consistent with the trace decay model.

As with experiment 2, the method used in experiments 4 and 5 allowed for simultaneous presentation of labelled shapes; there was thus no opportunity for the postulated trace decay to occur during the judgement task, yet systematic shifts in judgements again appeared. It thus appears that, while the decay model explains much of the data reported here, it is not in itself sufficient to account for the observed shifts in judgements when labels are paired with stimuli. The model needs to be systematically investigated before it can be conclusively evaluated, however there are suggestions in the data that instructions to subjects may be an important subsidiary determinant of the attention or weight assigned to the labels.

The results of experiment 4 suggested that covert labelling strategies may also be important determinants of judged similarity. The experiment set out to investigate the role of meaningfulness, as measured by association value, in the judgement task. There was a significant trend for certain pairs of nonsense shapes to show greater judged dissimilarity for subjects who saw one of each member of the pair as meaningful, and the other member as meaningless, than for subjects who saw both of that pair as either meaningful or as meaningless.

A series of multidimensional scalings of the data supported the interpretation that a covert labelling process was used by subjects such that when both shapes were interpreted as meaningful they both may have been given covert animistic labels, while a subjective interpretation of the shape as meaningless caused it to be given no such covert animistic label. These animism / nonanimism labels were postulated to have influenced similarity judgements by enhancing the similarity of pairs of shapes when both were given either animistic or nonanimistic labels, but reducing judged similarity for subjects who responded to one shape of the pair as animistic and the other as nonanimistic.

In experiments 4 and 5 overt labels were also paired with shapes in order to assess the weightings given to the two components of (a) stimulus, and (b) label when the components are synthesised into a judgement of overall labelled-stimulus similarity. The occurrence of adding, averaging and subtracting strategies was examined and all three were found to occur in the responses produced by individual subjects. However, there was little evidence that switching between the three strategies could be related to the magnitude of the differences in scale values between the component shape similarities and label similarities for any given pair of labelled stimuli.

The research discussed in this thesis indicated that there are small but consistent verbal factors which influence the judgement of stimulus similarities. It is not yet evident to what extent the effects shown could be considered to be special effects

in that they relate directly to the psychophysical method and the experimental situation employed. The finding in experiment 4 that covert labels influence similarity assessment is worthy of more detailed investigation, and the present results suggest interesting parallels with schema theory (Rankin, Markley and Evans, 1970). The trace decay model also needs further investigation; in its present form however, it is consistent with the position taken by Lenneberg (1967) that verbal factors influence recognition under certain experimental conditions - particularly when the task is difficult. Support for the trace decay model also comes from studies such as those of Bartlett (1932) and Kurtz and Hovland (1953) which showed that subjects who verbally categorise stimuli at the time of perceiving them exhibit superior retention of details of the stimuli when tested after a timed interval. The Carmichael, Hogan and Walter (1932) effect is also consistent with the trace decay model - these studies have been reported in detail in Chapter I.

A more detailed investigation of the effects of labels on the underlying dimensionality of stimuli as revealed by multi-dimensional scaling would also be useful. Analyses of experiments 3 and 4 of the present research tentatively suggest that when real-word labels are used to describe stimuli they do not add to the dimensionality of the stimuli but rather they assimilate along existing dimensions. This finding would be presumably related to the relevance of the labels used however.

Finally, it should be emphasised that the research reported

here investigates only the effect of simple verbal stimuli on relatively simple judgement processes. The role of language in more complex cognitive behaviour such as reasoning, theorising and other complex activities remains to be investigated.

---



REFERENCES

- AMERINE, M.A., PANGBORN, R.M., & ROESSLER, E.B. (1965)  
Principles of Sensory Evaluation of Food. Academic Press  
Inc., New York & London.
- ANDERSON, N.H. (1965 a) Averaging vs. Adding as a stimulus  
Combination Rule in Impression Formation. Journal of  
Experimental Psychology, 70, 394 - 400.
- ANDERSON, N.H. (1965 b) Primacy Effects in Personality Impression  
Formation Using a Generalised Order Effect Paradigm.  
Journal of Personality and Social Psychology, 2, 1 - 9.
- ANDERSON, N.H. (1966) Component Ratings in Impression Formation.  
Psychonomic Science, 6, 279 - 280.
- ANDERSON, N.H. (1967) Averaging Model Analysis of Set-size Effect  
in Impression Formation. Journal of Experimental  
Psychology, 75, 158 - 165.
- ANDERSON, N.H. (1968) A simple model for Information Integration.  
Chapter 73 in Abelson, R.P., Aronson, E., McGuire, W.J.,  
Newcomb, T.M., Rosenberg, M.J., and Tannenbaum, P.H. (Eds.)  
Theories of Cognitive Consistency: A Sourcebook. Chicago:  
Rand McNally (1968).
- ANDERSON, N.H. (1970) Integration Theory and Attitude Change.  
Technical Report No. 12. Center for Human Information  
Processing, Department of Psychology, University of  
California.
- ANDERSON, T.W. (1958) Introduction to Multivariate Statistical

Analysis. J. Wiley and Sons.

- ANDERSON, N.H. & JACOBSON, A. (1965) Effects of Stimulus Inconsistency and Discounting Instructions in Personality Impression Formation. Journal of Personality and Social Psychology, 2, 531 - 539.
- ARNOULT, M.D. (1953) Transfer of Predifferentiation Training in Simple and Multiple Shape Discrimination. Journal of Experimental Psychology, 45, 401 - 409.
- BARTLETT, F.C. (1932) Remembering: A study in Experimental and Social Psychology. New York, Macmillan, &
- BATTIG, W.F. (1956) Transfer from verbal pretraining to motor performance as a function of motor task complexity. Journal of Experimental Psychology, 51, 371 - 378.
- BEALS, R., KRANTZ, D.H. & TVERSKY, A. (1968) Foundations of multidimensional scaling. Psychological Review, 75, 127 - 142.
- BEARE, A.C. (1963) Color-name as a function of wavelength. American Journal of Psychology, 76, 248 - 256.
- BEEBE-CENTER, J.G. & WADDELL, D. (1948) A General Psychological Scale of Taste. Journal of Psychology, 26; 517 - 524.
- BERLYNE, D.E. (1965) Structure and Direction in Thinking. New York, Wiley.
- BOCK, R.D., & JONES, L.V. (1968) The Measurement and prediction of Judgement and Choice. Holden-Day.
- BROADBENT, D.E. (1958) Perception and Communication. New York, Pergamon Press.

- BROWN, J. (1958) Some Tests of the Decay Theory of Immediate Memory. Quarterly Journal of Experimental Psychology, 10, 12 - 21.
- BROWN, R. (1958) Words and Things. The Free Press, Glencoe, Illinois.
- BROWN, B.R. & DANSEREAU, D.F. (1970) Functional equivalence between same-different classifications and judged similarity of Markov patterns. Perception and Psychophysics, 7, 307 - 310.
- BROWN, R.W. & LENNEBERG, E.H. (1954) A study in language and cognition. Journal of Abnormal & Social Psychology, 49, 454 - 462.
- BRUNER, J.S., GOODNOW, J.J. & AUSTIN, G.A. (1956) A Study of thinking. New York, John Wiley & Sons.
- BRUNER, J.S., BUSIEK, R.D. & MINTURN, A. (1952) Assimilation in the immediate reproduction of visually perceived figures. Journal of Experimental Psychology, 44, 151 - 155.
- BUGELSKI, B.R. (1962) Presentation time, total time, and mediation in paired associate learning. Journal of Experimental Psychology, 63, 409 - 412.
- BURNHAM, R.W. & CLARK, J.R. (1955) A test of true memory. Journal of Applied Psychology, 39, 164 - 172.
- CANTOR, J.H. (1955) Amount of pretraining as a factor in stimulus predifferentiation and performance set. Journal of Experimental Psychology, 50, 180 - 184.
- CAPEHART, J., TEMPONE, V.J. & HEBERT, J. (1969) A theory of

stimulus equivalence. Psychological Review, 76, 405 - 418.

CAREY, J.E. & GOSS, A.E. (1957) The role of mediating verbal responses in the conceptual sorting behaviour of children. Journal of Genetic Psychology, 90, 69 - 74.

CARMICHAEL, L., HOGAN, H.P. & WALTER, A.A. (1932) An experimental study of the effect of language on the reproduction of visually perceived forms. Journal of Experimental Psychology 15, 73 - 86.

CARROLL, J.B. (1964) Language and thought. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

CHOMSKY, N. (1967) Review of Skinner's verbal behavior in Jakobovits, L.A. & Miron, M.S. Readings in the psychology of language. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

CHURCH, A.A. (1958) Introduction to mathematical logic. Princeton: Princeton University Press.

COFER, C.N. & FOLEY, J.P. (1942) Mediated generalization and the interpretation of verbal behaviour: I. Prolegomena. Psychological Review, 49, 513 - 540.

COFER, C.N. & MUSGRAVE, B.S. (1961) Verbal Behavior and Learning - Problems and Processes. New York, McGraw-Hill.

COOMBS, C.H. (1964) A theory of data. New York, Wiley.

DENO, S.L. & JENKINS, I.J. (1967) Semantic generalization of a voluntary response: Effects of responding in training and rate of presentation. Journal of Verbal Learning and Verbal

Behaviour, 6, 300 - 302.

DE RIVIERA, J. (1959) Some conditions governing the use of the cue-producing response as an explanatory device. Journal of Experimental Psychology, 57, 299 - 304.

DIXON, W.J. (1964) B M D computer programs manual. University of California & Los Angeles.

DOLLARD, J. & MILLER, N.E. (1950) Personality and Psychotherapy; - an analysis in terms of learning, thinking, and culture. New York, McGraw-Hill.

DUNCAN, C.P. (1953) Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. Journal of Experimental Psychology, 45, 1 - 11.

ELLIS, H.C. (1968) Transfer of stimulus predifferentiation to shape recognition and identification learning: Role of properties of verbal labels. Journal of Experimental Psychology, 78, 401 - 409.

EVANS, S.H. (1967) A brief statement of Schema theory. Psychonomic Science, 8, 87 - 88.

FEATHER, B.W. (1965) Semantic generalization of classically conditioned responses: A review. Psychological Bulletin, 63, 425 - 441.

FENN, J.D. & GOSS, A.E. (1957) The role of mediating verbal responses in the conceptual sorting behaviour of normals and paranoid schizophrenics. Journal of Genetic Psychology, 90, 59 - 67.

GARNER, W.R. (1962) Uncertainty and Structure as psychological

concepts. Wiley, New York.

GIBSON, E.J. (1940) A systematic application of the concepts of generalization and differentiation to verbal learning.

Psychological Review, 47, 196 - 229.

GIBSON, J.J. (1966) The senses considered as perceptual systems. Houghton Mifflin, New York.

GLAZE, J.A. (1928) The association value of nonsense syllables. Journal of Genetic Psychology, 35, 255 - 267.

GOSS, A.E. (1953) Transfer as a function of type and amount of preliminary experience with taste stimuli. Journal of Experimental Psychology, 45, 327 - 333.

GOSS, A.E. (1961) Verbal mediating responses and concept formation. Psychological Review, 68, 248 - 274.

GREENSPOON, J. (1955) The reinforcing effect of two spoken sounds on the frequency of two responses. American Journal of Psychology, 68, 409 - 416.

GREGSON, R.A.M. (1964) Modification of perceived relative intensities of acid tastes by ambient illumination changes. Australian Journal of Psychology, 16, 190 - 199.

GREGSON, R.A.M. (1968) A comparison of some models representing taste-mixture similarities in a cross-modal matching task. Supplementary paper prepared for the International Symposium on Sensory Evaluation of Food: Principles and Methods. Göteborg, Sweden.

- GREGSON, R.A.M. (1969) Determination of a gustatory L70 point for Sodium Benzoate. British Journal of Psychology, 60, 187 - 197.
- GREGSON, R.A.M. & SIMMONDS, M.B. (1971) Qualitative gustatory characteristics of disodium 5' guanylate. British Journal of Psychology. (In press.)
- HANAWALT, N.G. & DEMAREST, I.H. (1939) The effect of verbal suggestion in the recall period upon the reproduction of visually perceived forms. Journal of Experimental Psychology. 25, 159 - 174.
- HERMAN, D.T., LAWLESS, R.H. & MARSHALL, R.W. (1957) Variables in the effect of language on the reproduction of visually perceived forms. Perception and Motor Skills, 7, 171 - 186.
- HOBEN, T. (1968) Spatial models and multidimensional scaling of random shapes. American Journal of Psychology, 81, 551 - 558.
- HÖIJER, B. (1969) Mechanisms of Cognitive judgements. 67th report, Department of Psychology, University of Uppsala, Sweden.
- HÖIJER, B. (1970) On the consistency of similarity judgements. 77th report. Department of Psychology, University of Uppsala, Sweden.
- HOLTON, R.B. & GOSS, A.E. (1956) Transfer to a discriminative motor task as a function of amount and type of preliminary verbalization. Journal of General Psychology, 55, 117 - 126.

- HOPE, K. (1968) Methods of Multivariate Analysis. University of London Press Ltd., London.
- HORTON, D.L. & KJELDERGAARD, P.M. (1961) An experimental analysis of associative factors in mediated generalization. Psychological Monographs, 75, 11.
- HULL, C.L. (1939) The problem of stimulus equivalence in behaviour theory. Psychological Review, 46, 9 - 30.
- HULL, C.L. (1943) Principles of behavior. Appleton-Century Co. Inc., New York and London.
- HUNT, E.B. (1962) Concept learning: an information processing problem. Wiley, New York.
- HYMAN, R., & WELL, A. (1967) Judgements of similarity and spatial models. Perception and Psychophysics, 2, 233 - 249.
- HYMAN, R., & WELL, A. (1968) Perceptual Separability and spatial models. Perception and Psychophysics, 3, 161 - 165.
- JEFFREY, W.E. (1953) The effects of verbal and non-verbal responses in mediating an instrumental act. Journal of Experimental Psychology, 45, 327 - 333.
- JENKINS, J.J. (1955) Associative processes in verbal behavior: A report of the Minnesota Conference. University of Minnesota.
- JENKINS, J.J. (1961) Verbal Behavior and Learning - Problems and processes. McGraw-Hill, New York.
- JOHNSON, P.J. (1967) Nature of mediational responses in Concept Identification problems. Journal of Experimental Psychology, 73, 391 - 393.
- KAISER, H.F. (1958) The varimax criterion for analytic rotation in



- factor analysis. Psychometrika, 23, 187 - 200.
- KATZ, P. (1963) Effects of labels on children's perception and discrimination learning. Journal of Experimental Psychology, 66, 423 - 428.
- KATZ, P.A. & ZIGLER, E. (1969) Effects of labels on perceptual transfer: Stimulus and developmental factors. Journal of Experimental Psychology, 80, 73 - 77.
- KENDLER, T.S. (1961) Concept Formation. Annual Review of Psychology, 12, 447 - 472.
- KLAHR, D. (1969) A Monte-Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. Psychometrika, 34, 319 - 330.
- KLEMMER, E.T. & SHRIMPTON, M.W. (1963) Preference scaling via a modification of Shepard's proximity analysis method. Human Factors, 5, 163 - 168.
- KOPLIN, J.H. (1967) Mediated generalization of a motor response: All stages of the paradigm acquired in the laboratory. Journal of Verbal Learning and Verbal Behaviour, 6, 408 - 414.
- KRUSKAL, J.B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1 - 27.
- KURTZ, K.H. & HOVLAND, C.I. (1953) The effect of verbalisation during observation of stimulus objects upon accuracy of recognition and recall. Journal of Experimental Psychology, 45, 157 - 164.
- LACEY, H. & GOSS, A.E. (1959) Conceptual block sorting as a function

- number, pattern of assignment and strength of labelling responses. Journal of Genetic Psychology, 94, 221 - 232.
- LANTZ, D., & STEFFLRE, V. (1964) Language and cognition revisited. Journal of Abnormal and Social Psychology, 69, 472 - 481.
- LEFTWICH, J., NAWAS, M.M. & SIEGEL, S. (1969) Awareness in verbal conditioning: Objective assessment and automatic strengthening. Psychological Reports, 25, 51 - 59.
- LENNEBERG, E.H. (1967) Biological Foundations of Language. John Wiley and Sons Inc., New York.
- LUMSDAINE, A.A. (1939) Conditioned eyelid responses as mediating generalized conditioned finger reactions. Psychological Bulletin, 36, 650 (Abstract).
- MALLOY, T.E. & ELLIS, H.C. (1970) Attention and cue-producing responses in response mediated stimulus generalization. Journal of Experimental Psychology, 83, 191 - 200.
- MALTZMAN, I. & BELLONI, M. (1964) Three studies of semantic generalization. Journal of Verbal Learning and Verbal Behaviour, 3, 231 - 235.
- MALTZMAN, I. (1968) Theoretical conceptions of semantic conditioning and generalization. In T.R. Dixon & D.L. Horton (Eds.) Verbal behavior and General behavior theory. Prentice-Hall, Englewood Cliffs, New Jersey.
- MALTZMAN, I., LANGDON, B., & FEENEY, D. (1970) Semantic Generalization without prior conditioning. Journal of Experimental Psychology, 83, 73 - 75.
- MANDLER, G. (1961) Verbal behavior and learning - problems and

and processes. McGraw-Hill, New York.

McALLISTER, D.E. (1953) The effects of various kinds of relevant verbal pretraining on subsequent motor performance. Journal of Experimental Psychology, 46, 329 - 336.

McCORMACK, P.D. (1958) Negative transfer in motor performance following a critical amount of verbal pretraining. Perceptual and Motor Skills, 8, 27 - 31.

McGUIGAN, F.J. (1966) Thinking: Studies of Covert Language processes. Appleton-Century-Crofts, New York.

McKEE, J.P. & RILEY, D.A. (1962) Auditory transposition in 6 year old children. Child Development, 33, 469 - 476.

MELTON, A.W. (1963) Implications of short-term memory for general theory of memory. Journal of Verbal Learning and Verbal Behaviour, 2, 1 - 21.

MILLER, N.E. & DOLLARD, J. (1945) Social Learning and Imitation. K. Paul, Trench, Trubner & Co. Ltd., London.

MINK, W.D. (1957) Semantic generalization as related to word association. Ph.D. Thesis, University of Minnesota.

MITCHELL, M.J. & GREGSON, R.A.M. (1968) Interrelations of perceptual reports of smell, taste and irritance over the near-threshold range, for members of the N-aliphatic monohydric alcohol series. Perception and Psychophysics, 4, 13 - 18.

MOWRER, O.H. (1954) The Psychologist looks at language. American Psychologist, 9, 660 - 694.

MURDOCK, B.B. (1958) Effects of task difficulty, stimulus similarity and type of response on stimulus predifferentiation. Journal

of Experimental Psychology, 55, 167 - 172.

NAMIKAS, G. (1967) Concept identification as a function of relevance of pretraining and percentage of informative feedback. Psychonomic Science, 8, 261 - 262.

NEWELL, A. (1968) Judgement and its representation: an Introduction. Chapter 1 of Kleinmuntz, B. (Ed.) Formal representation of Human Judgement, John Wiley & Sons, Inc., New York.

NORCROSS, K.J. (1958) Effects of discrimination performance on similarity of previously acquired stimulus names. Journal of Experimental Psychology, 56, 305 - 309.

NURMINEN, A.M. (1965) Evaluation of a technique for measuring subjective similarity. Scandinavian Journal of Psychology, 6, 209 - 219.

OSGOOD, C.E. (1952) The nature and measurement of meaning. Psychological Bulletin, 49, 197 - 237.

OSGOOD, C.E. (1953) Method and theory in experimental Psychology, New York, Oxford University Press.

OSGOOD, C.E., SUCI, G.J., & TANNENBAUM, P.H. (1957) The Measurement of Meaning. University of Illinois Press, Urbana.

PAIVIO, A., & YUILLE, J.C. (1967) Mediation instructions and word attributes in paired-associate learning. Psychonomic Science, 8, 65 - 66.

PELTON, L.H. (1969) Mediational construction vs. mediational perception in paired associate learning. Psychonomic Science

17, 220 - 221.

PERYAM, D.R. (1960) The variable taste perception of sodium benzoate. Food Technology, 14, 383 - 386.

PFAFFMANN, C. (1962) In Koch, S. (Ed.) Psychology: A Study of a Science. McGraw-Hill Book Company Inc.

PIAGET, J. (1951) The origin of names. In Dennis, W. (Ed.) Readings in Child Psychology. Prentice-Hall, New York.

POREBSKI, O.R. (1966) Discriminatory and canonical analysis of technical college data. British Journal of Mathematical and Statistical Psychology, 19, 215 - 236.

POSNER, M.I., GOLDSMITH, R., & WELTON, K.E. (1967) Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology, 73, 28 - 38.

PRENTICE, W.C.H. (1954) Visual recognition of verbally labeled figures. American Journal of Psychology, 67, 315 - 320.

RANKIN, W.C., MARKLEY, R.P. & EVANS, S.H. (1970) Pythagorean distance and the judged similarity of schematic stimuli. Perception and Psychophysics, 7, 103 - 107.

RAZRAN, G. (1949) Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. Journal of Experimental Psychology, 39, 642 - 652.

REESE, H.W. (1968) The perception of stimulus relations: Discrimination learning and transposition. Academic Press, New York and London.

ROBINSON, J.O. (1970) The misuse of taste names by untrained Observers. British Journal of Psychology, 61, 375 - 378.

- ROBINSON, J.G. (1955) The effect of learning verbal labels for stimuli on their later discrimination. Journal of Experimental Psychology, 49, 112 - 114.
- ROSSMAN, I.L. & GOSS, A.E. (1951) The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. Journal of Experimental Psychology, 42, 173 - 182.
- RUSSELL, W.A. & STORMS, L.H. (1955) Implicit verbal chaining in paired-associate learning. Journal of Experimental Psychology, 49, 267 - 293.
- RUSSELL, W.A. (1955) Bi-directional effects in word association. In Associative processes in verbal behavior: A report of the Minnesota Conference. University of Minnesota, 1955.
- RUSSELL, W.A. (1961) Purpose and the problem of Associative Selectivity. Chapter 7, Cofer, C.N. and Musgrave, B.S. Verbal behavior and learning - Problems and processes. McGraw-Hill, New York.
- SACKS & RUSSELL, W.A. (1953) In Cofer, C.N. and Musgrave, B.S. Verbal behavior and learning - Problems and processes. McGraw-Hill, New York.
- SCHWARTZ, M. (1969) Instructions to mediate, recall time, and type of paired-associate list. Journal of Experimental Psychology, 81, 398 - 400.
- SHELLY, M.W. & BRYAN, G.L. (Eds.) (1964) Human Judgements and Optimality, New York. John Wiley & Sons.
- SHEPARD, R.N. (1957) Stimulus and response generalization: A

- Stochastic model relating generalization to distance in psychological space. Psychometrika, 22, 325 - 345.
- SHEPARD, R.N., HOVLAND, E.I. & JENKINS, H.M. (1961) Learning and memorization of classifications. Psychological Monographs, 75, No. 13, 1 - 42.
- SHEPARD, R.N. (1962) The analysis of proximities: Multidimensional scaling with an unknown distance function: I Psychometrika, 27, 125 - 140.
- SHEPARD, R.N. (1962) The analysis of proximities: Multidimensional scaling with an unknown distance function: II. Psychometrika, 27, 219 - 246.
- SHIPLEY, W.C. (1933) An apparent transfer of conditioning. Journal of General Psychology, 8, 382 - 391.
- SHIPLEY, W.C. (1935) Indirect Conditioning. Journal of General Psychology, 12, 337 - 357.
- SJOBERG, (1969) The dimensionality of similarity judgements. American Journal of Psychology, 82, 441 - 456.
- SKINNER, B.F. (1957) Verbal Behavior. Appleton-Century-Crofts Inc., New York.
- SPERLING, G. (1963) A model for visual memory tasks. Human Factors, 5, 19 - 31.
- SPIKER, C.C. (1956) Stimulus pretraining and subsequent performance in the delayed reaction experiment. Journal of Experimental Psychology, 52, 107 - 111.
- SPIKER, C.C. (1959) Performance on a difficult discrimination following pretraining with distinctive stimuli. Child

Development, 30, 513 - 521.

STAATS, A.W. (1968) Learning, Language and Cognition. Holt, Rinehart & Winston Inc. New York.

STAATS, C.K. & STAATS, A.W. (1957) Meaning established by classical conditioning. Journal of Experimental Psychology, 54, 74 - 80.

TURQUIN, P. (1969) Apprentissage verbaux et activités intellectuelles. Enfance, 5, 365 - 382.

UZNADZE, D.N. (1966) The Psychology of Set. Consultants Bureau, New York.

VANDERPLAS, J.M. & GARVIN, E.A. (1959) The association value of random shapes. Journal of Experimental Psychology, 57, 147 - 154.

VOSS, J.F. (Ed.) (1969) Approaches to Thought: A Symposium of the Learning Research and Development Center, University of Pittsburgh. Charles E. Merrill, Publishing Company, Columbus, Ohio.

WALLACH, M.A. (1958) On Psychological Similarity. Psychological Review, 65, 103 - 116.

WILSON, G.D., & GREGSON, R.A.M. (1967) Effects of illumination on perceived intensity of acid tastes. Australian Journal of Psychology, 19, 69 - 73.

YARMEY, A.D., & CSAPO, K.G. (1968) Imaginal and verbal mediation instructions and stimulus attributes in paired-association learning. Psychological Record, 18, 191 - 199.

YOSHIDA, M., & SAITO, S. (1969) Multidimensional scaling of the



taste of amino acids.(In Press, 1969)

YOUNG, F.W. & TORGERSON, W.S. (1967) Computer Program Abstracts - TORSCA, A Fortran IV Program for Shepard-Kruskal Multi-dimensional Scaling Analysis (CPA 259) Behavioral Science 12, 498.

YOUNG, F.W. (1968) A FORTRAN IV Program for Nonmetric Multidimensional Scaling. The L.L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina.

YUILLE, J.C. & PAIVIO, A. (1968) Imagery and verbal mediation instructions in paired-associate learning. Journal of Experimental Psychology, 78, 436 - 441.

